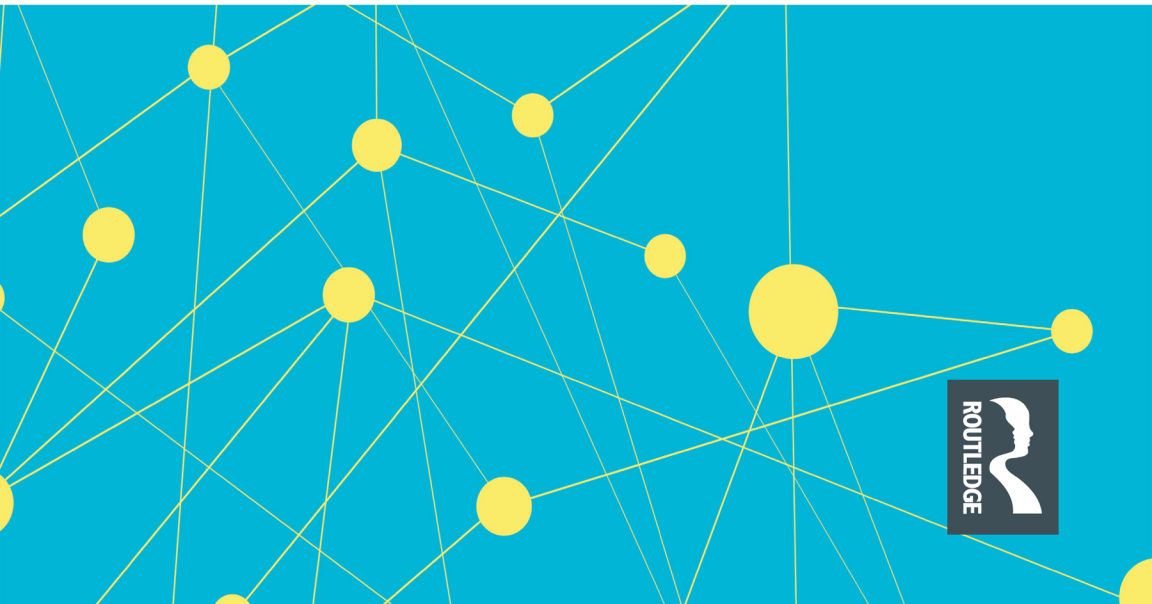
Routledge Festschrifts in Philosophy

WITTGENSTEIN AND BEYOND

ESSAYS IN HONOUR OF HANS-JOHANN GLOCK

Edited by
Christoph C. Pfisterer, Nicole Rathgeb, and
Eva Schmidt



Wittgenstein and Beyond

This volume celebrates the work of Hans-Johann Glock, a philosopher renowned for both his exegesis of Wittgenstein and his many contributions to debates in contemporary philosophy. It brings together 16 new essays by up-and-coming and distinguished philosophers engaging with Glock's work, and it concludes with a "Reflections and Replies" chapter in which Glock responds to his interlocutors.

Glock's distinctive philosophical voice features a rare combination of a Wittgenstein-inspired approach with a willingness to break away from Wittgenstein to tackle problems in an open-minded manner. The broad selection of essays included in this volume reflects Glock's wide-ranging philosophical interests and demonstrates the potential of applying Wittgensteinian insights to advance current systematic debates in philosophy. The chapters discuss Wittgenstein's philosophy, metaphilosophy, truth and language, animal minds and agency, reasons, and normativity.

Wittgenstein and Beyond will appeal to scholars and advanced students working on Wittgenstein, metaphilosophy, philosophy of mind, and philosophy of language.

Christoph C. Pfisterer is a postdoctoral teaching and research assistant at the University of Zurich. He works on various topics of early analytic philosophy, as well as on contemporary philosophy of perception. He has recently published on Wittgenstein and Frege, and he is currently completing a book manuscript on the language of perception.

Nicole Rathgeb is a postdoctoral researcher at the University of Bern and Visiting Research Fellow at the University of Hertfordshire. Her main research interests are in the philosophy of mind (belief, self-knowledge, and first-person authority) and philosophical methodology (conceptual analysis, conceptual engineering, and ordinary language philosophy). She is currently co-editing the *Metzler Handbook of Philosophy of Mind* and writing a book on conceptual analysis.

Eva Schmidt is Assistant Professor of Theoretical Philosophy at TU Dortmund. She works on epistemic reasons and reasons for action, explainable artificial intelligence, and perception. She is the author of *Modest Nonconceptualism: Epistemology, Phenomenology, and Content* (Springer, 2015); "Where Reasons and Reasoning Come Apart" (2021); and a co-author of "From Responsibility to Reason-Giving Explainable Artificial Intelligence" (2022).

Routledge Festschrifts in Philosophy

Mind, Language and Morality
Essays in Honor of Mark Platts
Edited by Gustavo Ortiz Millán and Juan Antonio Cruz Parcero

Sensations, Thoughts, Language
Essays in Honor of Brian Loar
Edited by Arthur Sullivan

Common Sense Metaphysics
Essays in Honor of Lynne Rudder Baker
Edited by Luis R. G. Oliveira and Kevin Corcoran

Wittgenstein and Beyond
Essays in Honour of Hans-Johann Glock
Edited by Christoph C. Pfisterer, Nicole Rathgeb, and Eva Schmidt

Teleological Structures in Human Life
Essays for Anselm W. Müller
Edited by Christian Kietzmann

For more information about this series, please visit: https://www.routledge.com/Routledge-Festschrifts-in-Philosophy/book-series/RFSP

Wittgenstein and Beyond

Essays in Honour of Hans-Johann Glock

Edited by
Christoph C. Pfisterer,
Nicole Rathgeb, and Eva Schmidt

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2023
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

ISBN: 978-1-032-05702-6 (hbk)
ISBN: 978-1-032-06587-8 (pbk)
ISBN: 978-1-003-20292-9 (ebk)

DOI: 10.4324/9781003202929

Typeset in Sabon
by codeMantra

Contents

Figures

Contributors

Maria Alvarez is Professor of Philosophy at King's College London. She is the author of *Kinds of Reasons; an Essay in the Philosophy of Action* and works mainly on philosophy of action, reasons, and normativity and is currently working on a monograph on agency, choice, and moral responsibility. She is also a co-editor of *Philosophy*.

Ansgar Beckermann is a Professor Emeritus at Bielefeld University, Germany. He held teaching positions in Osnabrück, Göttingen, Mannheim, and Bielefeld, and he has served as president for the German Society for Analytical Philosophy for several years. His research is mainly concerned with topics in the philosophy of mind, epistemology, philosophy of free will, and philosophy of religion. His most important publications include *Gründe und Ursachen* (1977), *Descartes' metaphysischer Beweis für den Dualismus* (1986), *Gehirn, Ich, Freiheit* (2008), *Glaube* (2013), *Naturalismus* (2021), and *The Oxford Handbook of Philosophy of Mind* (2009); ed. with B. McLaughlin and S. Walter.

Hanoch Ben-Yami teaches philosophy at Central European University, Vienna. He has published on logic; philosophy of logic; philosophy of language; philosophy of mind and action; philosophy of physics, especially with relation to Special Relativity; medical ethics; Descartes and early modern thought; and Wittgenstein.

Gerhard Ernst holds a chair for philosophy at the Friedrich-Alexander-Universität Erlangen-Nürnberg (Germany). From 2018 till 2021, he was President of the Deutsche Gesellschaft für Philosophie (German Society for Philosophy). His main research areas are epistemology and metaethics. His works include *Das Problem des Wissens* (The Problem of Knowledge) (Paderborn 2002) and *Die Objektivität der Moral* (The Objectivity of Morality) (Paderborn 2008). For more details, see: http://www.philosophie.phil.uni-erlangen.de/lehrstuehle/-lehrstuhl-I/ernst.shtml.

Hans-Johann Glock is Professor of Philosophy at the University of Zurich and a Principal Investigator in the National Centre of Competence in

Research *Evolving Language*. His main research areas are philosophy of mind, philosophy of language, history of analytic philosophy, normativity of language, concepts, and animal minds. He is the author of *La Mente de los Animales: Problemas Conceptuales, What is Analytic Philosophy?*, *Quine and Davidson on Language*, *Thought and Reality*, and *A Wittgenstein Dictionary*. He is the editor of *Intentionality and Language*, *Wittgenstein: A Critical Reader*, and *The Rise of Analytic Philosophy* and a co-editor of *Wittgenstein's Philosophical Investigations: Text and Context*, *Wittgenstein and Quine*, *A Companion to Wittgenstein*, and *Wittgenstein and Analytic Philosophy – Essays for P. M. S. Hacker*.

Maja Griem is a PhD student at the Institute of Philosophy II at Ruhr-University Bochum and a member of the Research Training Group "Situated Cognition," funded by the Deutsche Forschungsgemeinschaft (DFG: GRK-2185/2). She is working on a project on animal cognition with a focus on empathy, social cognition, and scaffolded cognition. She completed two master programs, namely M.A. in Philosophy and M.Sc. in Cognitive Science.

Brad Hooker worked for Virginia Commonwealth University during 1986–1993 and the University of Reading during 1993–2019. He is best known for his book *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality* (OUP, 2000).

John Hyman is Grote Professor in the Philosophy of Mind and Logic at UCL. He was an Official Fellow and Tutor in Philosophy at The Queen's College, Oxford from 1988 to 2018, and Professor of Aesthetics in the University of Oxford from 2008 to 2018. He was a Getty Scholar at the Getty Research Institute, Los Angeles in 2001–2002, a Fellow at the Wissenschaftskolleg zu Berlin in 2002–2003, a Leverhulme Major Research Fellow in 2010–2012, and Professeur Invité at the Université Paris-Sorbonne in 2014–2015. He edited the *British Journal of Aesthetics* from 2008 to 2019, together with Elisabeth Schellekens. He joined the Philosophy Department at UCL in 2018, and he began the five-year ERC project *Roots of Responsbility* at the same time. His publications include *The Objective Eye* (Chicago University Press, 2006), *Action, Knowledge and Will* (Oxford University Press, 2015), and *The Blackwell Companion to Wittgenstein* (Wiley, 2017), which he edited jointly with Hans-Johann Glock.

Wolfgang Künne is Professor Emeritus of Philosophy at the University of Hamburg. He has held teaching positions at Bielefeld, Fribourg, Oslo, Heidelberg, Venice, Salzburg, and Reading. He gave the Gareth Evans Memorial Lecture at Oxford (1998), and in 1999, he was the Winchester Lecturer in Philosophy at the University of Oxford. In 2009, he was the first to win the "Frege-Award" of the German

Society for Analytic Philosophy. His research focuses on the philosophy of language, metaphysics, and the origins of analytic philosophy in Prague, Jena, and Cambridge. His most important publications include *Abstrakte Gegenstände* (1983, 2nd enlarged edition 2007), *Conceptions of Truth* (2003), *Essays on Bolzano* (2008), *Die Philosophische Logik Gottlob Freges* (2010), and *Epimenides und andere Lügner* (2013). He is currently working on a book on Bernard Bolzano's Life and Work to appear in four volumes in 2022/2023. He is a full member of the Göttingen Academy of Sciences and of the Academia Europaea.

Julia Langkau is an Assistant Professor at the University of Geneva (PRIMA grant, Swiss National Science Foundation SNSF). Previously, she was an SNSF research fellow at the University of Fribourg, a visiting research fellow at the University of Miami, a ZIF Marie Curie fellow at the University of Konstanz, and a postdoctoral researcher at the University of Zürich. Her research areas are the philosophy of mind, epistemology, and aesthetics.

Albert Newen is Full Professor of Philosophy with a focus on philosophy of mind, language, and cognition at the Ruhr-University Bochum (RUB). He is the director of the Interdisciplinary Center for Mind and Cognition at RUB and speaker of the Interdisciplinary Research Training Group on "Situated Cognition" (DFG-Graduiertenkolleg). Recent edition: Newen, de Bruin, Gallagher (eds): Oxford Handbook of 4E Cognition, Oxford: OUP (2018).

Christian Nimtz is Professor of Theoretical Philosophy at Bielefeld University, Germany. He works on topics in the philosophy of language, modal epistemology, metaphilosophy, and the philosophy of mind. His current research focuses on the meta-semantics of predicates, especially natural kind predicates, the prospects for two-dimensional semantics, the methodological import of conceptual analysis and conceptual engineering, and on related issues.

Simone Pika is head of the research group Comparative BioCognition at the Institute of Cognitive Science at the University of Osnabrück. Her research centers on gaining a better understanding of the evolution of language and cognition by specifically focusing on communicative and cognitive skills of nonhuman primates and corvids.

Constantine Sandis is Professor of Philosophy at the University of Hertfordshire, a Founding Director of Lex Academic, and a Fellow of the Royal Society of Arts. He is also the secretary of the British Wittgenstein Society and the editor of *Anthem Studies in Wittgenstein*. Sandis is the author of *The Things We Do and Why We Do Them* (Palgrave 2012) and *Character and Causation* (Routledge, 2018), as

well as the editor or co-editor of a dozen volumes. His forthcoming books include *How to Understand Others* (Yale University Press), *Wittgenstein's Lion* (Bloomsbury), and *Wittgenstein on Other Minds* (Anthem).

Severin Schroeder is Associate Professor of Philosophy at the University of Reading. He has written four monographs on Wittgenstein: *Wittgenstein: The Way Out of the Fly Bottle* (Polity, 2006), *Wittgenstein on Mathematics* (Routledge 2020), *Das Privatsprachen-Argument* (Schöningh/Mentis, 1998), and *Wittgenstein Lesen* (Frommann-Holzboog, 2009). He is the editor of *Wittgenstein and Contemporary Philosophy of Mind* (Palgrave 2001) and *Philosophy of Literature* (Wiley-Blackwell 2010).

Joachim Schulte recently retired from teaching at the University of Zurich, Switzerland. His publications include numerous articles and four books on Wittgenstein as well as critical editions of his main works; with P.M.S. Hacker, he edited and translated the revised fourth edition of Wittgenstein's *Philosophical Investigations* (2009).

Helen Steward is Professor of Philosophy of Mind and Action at the University of Leeds and is a Fellow of the British Academy. She is the author of *A Metaphysics for Freedom* and *The Ontology of Mind* (1997). Her interests lie mainly in the philosophy of action and free will, the philosophy of mind, and the metaphysical and ontological issues which bear on these areas. She is currently writing a book on causation.

Daniel Whiting is Professor of Philosophy at the University of Southampton. His publications include *The Range of Reasons* (Oxford University Press) and the edited volume *The Later Wittgenstein on Language* (Palgrave). He recently led the project *Higher-Order Evidence in Epistemology, Aesthetics, and Ethics*, supported by the Arts and Humanities Research Council.

Markus Wild is Professor of Theoretical Philosophy at the University of Basel. He completed his PhD in Basel on the mind of animals in early modern philosophy, and he habilitated on biosemantics at the Humboldt University Berlin. His areas of research are the philosophy of animal minds, animal ethics, philosophy of mind, and, more recently, the philosophy of Nietzsche.

Introduction

Christoph C. Pfisterer, Nicole Rathgeb, and Eva Schmidt

The present anthology is a festschrift and festschrifts are usually published no earlier than the occasion of the 60th birthday of well-known, influential, and seasoned intellectuals to honour their academic work. This collection celebrates the work of Hans-Johann Glock, and we still find it hard to believe that he has reached this milestone. Anyone who has ever dealt with Hanjo professionally or privately will know him as that lively, humorous, and thoughtful person who is always on his toes, never shies away from philosophical debate, never turns down an offer, provided others also benefit from his effort to make the impossible possible, and has an incredibly big heart in every respect. However, numbers do not lie, and even a hasty glance at his extensive list of publications allows the only reasonable conclusion that, for all his hustle and bustle, the man must have a few years under his belt to have achieved all this—happy birthday!

When we started work on the present volume, we were simply overwhelmed by the unanimously approving responses from Glock's former and current colleagues, mentors, students, and friends. They all agreed without hesitation to contribute to the festschrift, and it was a palpably emotional moment for all who, after nearly two years of the pandemic, gathered at a three-day symposium in September 2021 in Zurich to discuss the papers published here. The circle of intellectual friends reaches further than what could be fitted between two book covers. However, as editors, we are not only pleased by the many outstanding contributions written especially for this collection, but also deeply impressed by the broad scope of topics and by the dedication with which the authors engage with Glock's work. The selection of essays included in this volume reflects Glock's wide-ranging philosophical interests and demonstrates the potential of applying Wittgenstein's insights to advance current debates in philosophy. Glock's rare combination of a Wittgenstein-inspired approach with a willingness to break away from Wittgenstein to tackle problems in an open-minded way makes his a distinctive voice in contemporary philosophy.

Before giving an overview of the individual contributions, we will briefly sketch some stages in the jubilarian's academic career and touch

DOI: 10.4324/9781003202929-1

on the main areas of his philosophical research. Born and raised in the Black Forest in Germany, Glock initially planned to study physics and mathematics. His interest in philosophy was first sparked by a long-distance radio course *Praktische Philosophie/Ethik* by Karl-Otto Apel, and by a summer school of the German Academic Scholarship Foundation that he attended while still serving as a conscientious objector in the local hospital. There he was assigned to a group of young prize-winning mathematicians working on p-adic numbers. Soon he found himself gravitating towards another group of students discussing Aristotle's *Topoi*. To his delight, he discovered not only that philosophy is just as fascinating as mathematics, but also that he might be better at it. Glock went to study philosophy at Tübingen, where he would have had to choose between the right-wing Hegelians and the left-wing Hegelians, unless he wanted to risk being considered an intellectually backward Kantian. Due to his interest in natural science, he was more attracted to logical positivism and analytic philosophy, and had the good fortune to fall in with a dissident group of analytic philosophers who were secretly reading Tugendhat. His philosophical preferences were shaped since his school days by the physicist and philosopher Walter R. Fuchs, who characterises analytic philosophy as follows: '*a* [not the] pretty reasonable kind of philosophising, which is quite suited to the needs of a society shaped by *natural science* and *technology*' (Fuchs 1972, 10; *trans. HG*).

After passing his prelims, Glock received a stipend from the German Academic Exchange Service to go to the United Kingdom and was accepted as a visiting scholar at Oxford. Coming to Oxford at the tail end of its golden age, he met people like Michael Dummett, Peter Strawson, Derek Parfit, Jennifer Hornsby, Joseph Raz, Bernard Williams, John McDowell, and Peter Hacker. To Glock, this felt like a promotion from the Conference League to the Premier League. Although the debates were challenging and competitive, people talked to each other in a manner that was civilised and often very constructive. It was the time when the 'Davidson research programme' hit Oxford. Many people from professors down to BPhil students believed that the philosopher's stone consisted in resolving the issues between Dummett and Davidson over the shape of a systematic theory of meaning for natural languages. While some were desperately trying to fathom all the details of Dummett's notoriously difficult articles on what a theory of meaning is, others were busy trying to find out what the real arguments were in Davidson's seminal 'Actions, Reasons, and Causes'.

Glock's original plan was to work on transcendental arguments, but Strawson was unavailable for supervision in the first term. As a result, he was asked whether he would like to do something with a 'local flavour' and work on Wittgenstein with a certain Peter Hacker. And so it happened that in his first year at Oxford, he had weekly tutorials first with Hacker on Wittgenstein, in the second term with Strawson on Kant

and transcendental arguments, and in the final term with McDowell on Wittgensteinian themes. Next to Hacker, Strawson is the person and teacher who impressed and influenced him the most, not only because of his philosophical style, which was more conciliatory than that of other heavyweights. He also got to like him as a person and thought that Strawson was simply right (or almost right) on many issues; for example, reference, particulars, universals, categories, truth, and the proper role of formal methods in philosophy. At one time, he left his copy of Kant's first critique in his office after a supervision, and Strawson came running after him, waving the book and shouting: 'Mr. Glock, your *Critique of Pure Reason*—never leave without it!'

About Hacker, Glock says that he not only had more to teach about Wittgenstein than he could possibly have learned; he also jerked him out of his mainstream analytic thinking. Glock started out by presenting the essays on the syllabus, peddling all the orthodoxies of philosophy of language that he had imbibed through reading Tugendhat. But soon he came to realise that his elucidations would not pass muster with Hacker, who interrupted him at every juncture. Feeling like he was placed in an intellectual tumble dryer, on one occasion he argued eloquently that the meaning of a sentence consists in its truth conditions, and Hacker replied: '"The moon is blue" is true if and only if the moon is blue.—Well, what sort of condition is that?!' Working with Hacker meant not to flinch from questioning fashionable intellectual paradigms. After a term of feeling like a 'complete idiot'—Glock's own words—he was both surprised and flattered when he was asked to comment on drafts of the second volume of the commentary on Wittgenstein's *Philosophical Investigations* that Hacker was co-authoring with Gordon Baker.

Apart from his academic teachers, Glock learned a lot from his peers, some of whom are contributing to this volume. In his first year, he enjoyed stimulating discussions with David Bakhurst and Olav Gjelsvik. When he returned to Oxford as a DPhil student after completing his MA in Berlin with Tugendhat, the scene was even richer. John Hyman and Maria Alvarez became close philosophical interlocutors, and any remark over lunch could easily turn into a protracted philosophical debate. Hyman and Glock set up discussion and reading groups with a truly impressive list of members; people such as Peter Strawson, Joseph Raz, Anthony Kenny, Bede Rundle, Oswald Hanfling, Maria Alvarez, John Cottingham, Jonathan Dancy, and Hanoch Ben-Yami assembled in Glock's office at St. John's College to discuss philosophical papers.

After finishing his DPhil with Hacker on Wittgenstein's conception of philosophy, Glock had a junior research fellowship at St. John's College. Hacker explained to him that he would never get a fellowship at Oxford, since Hacker's references were like the 'kiss of death'. On his advice, Glock accepted a permanent position as a lecturer at Reading. This proved to be another stroke of luck in Glock's academic career. Under

the leadership of John Cottingham, the department combined intellectual vigour, competition, and *esprit de corps*. Glock became friends with Brad Hooker, and towards the end of his stay the department was joined by Severin Schroeder. This further strengthened its credentials, especially yet not exclusively as regards Wittgenstein. With the personnel changes at Oxford, Reading could moult into a new centre for Wittgenstein studies, which also included John Preston and Max de Gaynesford.

As inspiring as the intellectual environment was, the teaching load was high, and of all people Glock was assigned to hold the introductory lectures on Plato. In addition, there was immense pressure to publish in order to satisfy the research assessment exercise. At a time where the department was threatened with closure, Glock decided to accept the commission to write the *Wittgenstein Dictionary* (Glock 1996). About that time, Glock says: 'I could not have worked any harder if you had put a gun to my head'—especially because he and his wife Gabi Franz also had two wonderful daughters to raise. Glock remembers one of the traditional Christmas dinners with friends at their house. In between the courses that were served, he had to complete an entry of the *Dictionary*; and we challenge our readers to find out which entry that was.

Having been promoted to a professorship at Reading, the decision to leave for the University of Zurich was not an easy one. Family reasons, a love of the great outdoors, and the prospect of a new challenge prevailed. There were losses as well as gains, however. When Glock moved to the University of Zurich in 2006, he soon found himself embroiled in three controversies (see Glock 2012). The first was a xenophobic campaign against 'too many German professors' at Swiss universities. The second was his own campaign for abolishing the Latin requirement at the Faculty of Philosophy; in the eyes of some, this turned Glock into 'one of biggest threats to Western civilisation since Genghis Khan' (Glock's words). And the third was a lingering hostility to analytic philosophy, especially within the humanities and the educated public. Especially regarding that last point, Glock is happy to diagnose a sea-change. The University of Zurich is now a flourishing centre for broad-minded and historically informed analytic work in both theoretical and practical philosophy, where Glock not only finds himself surrounded by excellent colleagues, but also has been able to reconnect with old friends like Joachim Schulte or Katia Saporiti. Glock is also trying to make the most of the opportunities that the University affords for interdisciplinary collaboration with the life and cognitive sciences, e.g. as part of the ongoing National Centre of Competence in Research *Evolving Language*.

In terms of themes that have shaped his work in recent decades, Wittgenstein clearly remains an integral part and inspiration—the sceptic may be convinced by the bibliography printed at the end of this volume. Nevertheless, Glock is not following him faithfully on every point, as the title of this festschrift indicates. He thinks that we should hold on to

Wittgenstein's critique both of referential conceptions of meaning and of the Cartesian or 'inner-outer picture' of the mind, as well as to a distinction between philosophical and scientific questions. At the same time, we should relinquish all vestiges of the idea that being exercised by philosophical questions is a sign of some kind of intellectual disease; and we should also resist Wittgenstein's occasional anti- or irrational tendencies. There is no gainsaying the fact that Wittgenstein sold at least some of the tickets that the therapeutic interpreters and the so-called 'New Wittgensteinians' travel on. In Glock's view, by contrast, one can share a *critical* conception of philosophy as Kant and Strawson did, without regarding philosophical questions as symptoms of a disease. He recognises philosophy as a kind of meta-enterprise, not directly continuous with empirical or formal science, but engaged in conceptual clarification as required when addressing fundamental questions about thought and reality. In the same vein, he does not share Wittgenstein's *ab initio* rejection of systematic philosophising. Wittgenstein is certainly right that no standard definitions can be given for many philosophically important notions. Nevertheless, the attempt to reach definitions in terms of necessary and sufficient conditions, say for *meaning*, *intentional action*, *truth*, or *norm*, is always valuable and instructive, even when it fails.

Further points of divergence concern Wittgenstein's views on the semantics–pragmatics distinction, as well as his take on necessary truth and religion. According to Glock, Wittgenstein rightly recognised that the rules constitutive of word meaning and the sense of sentences interact with contextual features, but some of his claims seem to blur the distinction between semantics (lexical meaning) and pragmatics (communicative purpose and implicatures) entirely. Moreover, it is difficult to understand why he was so reticent about acknowledging that *a priori* necessary true propositions like those of logic and mathematics are truth-apt and can be known. Wittgenstein had a lifelong problem with simply accepting the fact that 'grammatical propositions', as he calls them, can be true and can be known. And finally, it is difficult to appreciate why Wittgenstein thinks that it is the rationalist critics of religion like the *Encyclopédistes*, Kant, or Russell that are conceptually muddled, rather than the true believers and the theological fideists like Pascal, or indeed Wittgenstein himself.

With respect to methodology, Glock is a proponent of conceptual analysis, to wit, *connective* (as opposed to reductive) analysis in the vein of Strawson (1992). With regard to concepts, Glock has defended a cognitivist view in a series of articles (2006, 2010b, 2010c, 2021). He claims that concepts are principles or rules for certain intellectual operations, in particular the operations of classification and inference. This definition sits well with the way in which the word 'concept' (and its counterparts in other languages) is used in logic, philosophy, psychology, and the history of ideas. Moreover, it can account for both the role concepts play

in the cognitive lives of individuals and the logical role they play as the components of propositions that enter into inferential relations.

Glock has coined a term for the specific version of conceptual analysis he favours: 'impure conceptual analysis' (see Glock 2013, 2017). Its impurity comes to the fore in connection with those philosophical topics which are also of interest to scientists—such as the topic of animal minds. With regard to the questions asked in such fields, for example, 'Do animals reason?', we can still in principle distinguish between, on the one hand, the philosophical-*cum*-conceptual question of what a creature has to be able to do in order for it to count as reasoning and, on the other hand, the empirical question whether or not (certain) animals do manifest the relevant behaviour. However, Glock argues, in a fertile investigation of topics such as animal cognition, conceptual and factual issues interact dynamically. It would be wrong to think of the conceptual side of things as purely *a priori*: Empirical findings can guide us in the analysis of complex and highly contested concepts such as that of reasoning; and they can contribute to establishing the inadequacy or barrenness of suggested conceptual explanations. What is more, it is not always clear where exactly the line between conceptual and factual questions is to be drawn—and the same holds for the line between conceptual and *methodological* questions. 'Morgan's Canon', for example, the principle that animal behaviour should not be explained in terms of capacities that are more demanding than necessary, is of utmost interest in the philosophical debate on animal minds, but is not reducible to a purely conceptual matter.

As evidenced by influential contributions (such as Glock 2000, 2009, 2010a), Glock has had a major impact on the philosophical debate over animal minds. However, he did not hit upon the topic through having pets. Rather, his interest is inherently theoretical, using the topic of animal minds as a starting point to get clearer on the nature of mind (is it representational?), concepts (and their relation to language), or intentional states (in what sense do they have a 'content'?) quite generally. A distinction that lies at the heart of Glock's thought about animal minds is that between differentialists/lingualists and assimilationists. The former tie mentality to language and consequentially argue that animals are not capable of thought; the latter regard the differences between humans and other animals as merely gradual, denying that there is a large gap between the mental capacities of humans and those of animals. Glock himself occupies a kind of middle ground between these two positions, arguing that animals can think without possessing concepts, and possess concepts without having language, but also acknowledging that non-human animals' lack of language (or concepts) significantly limits their mental capacities, and thereby the scope of what they can desire, believe, and know.

Some of Glock's most recent work has focused on the nature of reasons, normativity, and rationality. Glock has closely investigated the reasons for which we act, the normativity of reasons, and the nature of rational agency, both in connection with the question of animal agency (Glock 2019) and independently of that question (Glock 2014; Glock and Schmidt 2021). In an objectivist, anti-psychologist vein, Glock insists that the reasons for which we act are not our beliefs or desires, understood as mental states, but rather *what* we believe or desire—the facts or apparent facts (states of affairs) we believe, but also the goals we strive to achieve. He argues that such an objectivist picture can more easily accommodate the possibility that animals act for reasons. Since some non-human animals can act in pursuit of goals that they adopt for themselves, they can even be said to act rationally in a relatively demanding sense, or so Glock argues.

Each chapter of the festschrift deals with one of the topics from Glock's research areas as outlined above. The collection is divided into four parts, the first of which is devoted to the interpretation of Wittgenstein's philosophy.

In Chapter 1, Severin Schroeder discusses the idea, based on an interpretation of Wittgenstein's *On Certainty*, that there is a category of seemingly empirical propositions which are, in fact, grammatical propositions and exempt from doubt due to their fundamental role in our language. He sides with Glock in disputing this interpretation, arguing that these 'hinge propositions' are fallible empirical propositions after all. What is fundamental to our language game, Schroeder argues, are not specific propositions, but the standards of rationality that make us regard such propositions as certain, and those standards are reflected in our everyday judgements.

Joachim Schulte's contribution (Chapter 2) concerns Wittgenstein's attitude to history, and in particular the question whether Wittgenstein's approach to philosophy can be regarded as historicist. Schulte examines Glock's claim that Wittgenstein endorses a 'minimalist historicism', according to which knowledge of conceptual history is useful but not essential for philosophical insight. He clarifies Wittgenstein's use of some key terms (such as 'spirit' and 'culture') and offers a fruitful reading of some central passages in Wittgenstein's writings.

Daniel Whiting (Chapter 3) discusses the controversial notion of *nonsense* that figures prominently in Wittgenstein's work. According to one interpretation of the later Wittgenstein, nonsense can result from the improper combination of meaningful expressions; according to another, it can only ever result from privation—from a failure to assign meaning to one or more of the relevant expressions. Whiting takes issue with Glock's defence of the view that Wittgenstein allows for combinatorial nonsense and develops his own version of the privation view, arguing that, for

Wittgenstein, nonsense results from a failure to use an expression in a way that has a point.

Constantine Sandis's contribution (Chapter 4) discusses Wittgenstein's perplexing remark that if a lion could talk, we could not understand it. On the most charitable reading, according to Glock, Wittgenstein's point is not that we would be unable to understand a lion that spoke a human language, but that if lions had a feline language of growls and roars, we would be unable to learn it, since their form of life and behavioural repertoire are so different from our own. Sandis argues, however, that the issue is not what form a lion's language might take, but whether it is possible in principle for a human to come to understand a lion's use of language. To address this question, it is important to distinguish between understanding what the lion says and understanding the lion itself.

The second part of the festschrift is entitled 'Metaphilosophy, Truth, and Perception'. Wittgenstein remarked that what a person says or thinks is true if, and only if, things are as she says or thinks they are. In Chapter 5, Wolfgang Künne takes this truism as his starting point in the exposition of a definition of the predicate '*x* is true'. In doing so, he avails himself of quantification into the position of a full sentence and of the concept of a proposition. He defends this account (elaborated in Künne 2003) against the objection that it is necessary to invoke the notion of truth to explain sentential quantification and the notion of a proposition, and that the definition is therefore circular. Finally, he argues that a definition of the truth predicate suffices for an explanation of the concept of truth, since the meaning of the truth predicate is contained in the meaning of the sentence prefix 'It is true that ()'.

Ansgar Beckermann (Chapter 6) raises the question to what extent the method of 'impure conceptual analysis' that Glock endorses and his conception of a division of labour between philosophy and the empirical sciences can be applied to the philosophy of religion. Beckermann argues that this methodological picture does not quite fit the case of the philosophy of religion, where neither conceptual analysis nor the empirical sciences seem to play an important role. Rather, the method of this field of philosophy is to point to general facts and discuss their implications for questions such as 'Does God exist?'

The contribution by Christian Nimtz (Chapter 7) also takes a critical look at Glock's methodological approach. He argues that contrary to Glock's conception, the job of philosophy is not confined to the elucidation of concepts and the theoretical assessment of scientific theories: Philosophy can also procure evidence suited to support empirical scientific hypotheses. By way of an in-depth case study of John Perry's 'The Essential Indexical', Nimtz advocates the view that one important contribution of philosophy and, particularly, philosophical thought experiments consists in the acquisition of 'known near-actual truths'.

These are propositions that we know could easily be true, and they can abductively support empirical theories.

In his contribution (Chapter 8), John Hyman traces the development of the philosophy of perception in the twentieth century away from expressly empiricist theories, such as phenomenalism and Lockean indirect realism, towards the causal theory of perception and the disjunctivism of that theory's critics. He examines Strawson's argument in favour of the causal theory and assesses Snowdon's objection to it. He then presses this objection further, bringing it to bear against Snowdon's own disjunctivist account as well, and draws the conclusion that the disjunctivist's retreat from empiricism has not gone far enough.

Part III of the festschrift engages with Glock's work on animal minds. Markus Wild (Chapter 9) examines what it takes for non-human animals to possess conative and cognitive capacities and how researchers should proceed in order to determine the contents of animal mental states. He argues that both questions can only be answered on the basis of an augmentation of Glock's explanatory framework by what Wild calls the 'teleosemantic capacity approach' and 'hydraulic ethology'. This augmentation is also needed in order for Glock's 'master argument for animal cognition' to succeed and in order to dispose of a number of false dichotomies.

The chapter contributed by Maria Alvarez (Chapter 10) critically examines Glock's defence of the thesis that animals can act for reasons. She starts by commenting on the distinction between the 'subjectivist' and the 'objectivist' conception of reasons. Then, she raises two objections against Glock's central argument. She argues, first, that the forms of animal behaviour that Glock appeals to in his argument can be explained in a way that does not grant awareness of facts to animals, that is, awareness *that things are thus-and-so*. Secondly, Glock underestimates the complexity of the capacities that would have to be ascribed to animals on his preferred explanation of their behaviour.

The chapter by Albert Newen, Maja Griem, and Simone Pika (Chapter 11) focuses on empathy in human and non-human animals. It is written in support of Glock's endeavour to change our anthropological views and thereby pave the way for a better understanding of the cognitive abilities of animals. Its topic is empathy and the extent to which the importance of empathy for humans is anchored in evolution. In order to answer this question, Newen, Griem, and Pika set out a new conceptual framework in which different stages in the ontogenetic development of empathy can be described. The authors show further how this framework can be applied to assess empathy across different species of animals.

Helen Steward (Chapter 12) approaches Glock's work on animal minds from a more assimilationist rather than a more differentialist perspective (by contrast to Alvarez and Ben-Yami). She investigates whether

Glock may be guilty of unjustified zoocentrism in his denial of agency—and even behaviour—to plants as opposed to animals. Glock regards plants as mere information-processers. Steward argues, however, that this verdict cannot be justified given the way in which he describes and illustrates the difference between information processing on the one hand and behaviour on the other. Moreover, if Glock's further distinction between mere behaviour and agency is spelt out in a way that is consistent with his other views, he might even have to grant the possibility that some plants are agents.

While Glock's work has drawn attention to the intelligence of some animals and the respects in which animal mentality and agency sometimes are continuous with our own, Hanoch Ben-Yami's contribution (Chapter 13) is an attempt to identify which mental capacities separate us from animals. He identifies the command of logical concepts as a significant difference. He then shows how many behavioural, intellectual, emotional, and moral capacities depend on this mastery and examines recent empirical research into the limitations of intelligent animals in these respects.

In Chapter 14, Julia Langkau argues that current approaches to creativity blend together two different notions of creativity which should be kept apart: product creativity and process creativity. If we distinguish these two notions, we can resolve some apparent conceptual tensions concerning creativity and better explain the sense in which exceptional humans, animals, artificial intelligence generated art and inventions, and children's drawings can each be called 'creative'.

Part IV of the festschrift centres on the topics of normativity and reasons. In three sections, devoted to the topics of rationality, reasons, and rules, respectively, Brad Hooker engages with various themes from Glock's work (Chapter 15). The chapter begins with a discussion of four conceptions of rationality identified by Glock. Hooker explores how they are to be understood and what they each entail on a plausible reading. Proceeding from the conception of rationality as responsiveness to reasons, he explores the idea of reasons 'out-weighing' one another and assesses the extent to which different moral theories are able to accommodate the distinction between '*pro tanto*' and 'all things considered' oughts. In the final section of the chapter, Hooker spells out some connections between reasons and rules and argues that neither of the two categories can be reduced to the other.

Gerhard Ernst's contribution (Chapter 16) is concerned with reasons for emotions and the question under what circumstances emotions can be regarded as rational or irrational. On the basis of seven short case studies, he delineates various respects in which emotions (and specifically fear) may be irrational. Along the way, he expounds how two different conceptions of rationality (one according to which rationality is a

matter of consistency among mental states, and one according to which it is a matter of correctly responding to reasons) can both shed light on the way the concept applies to emotions.

The book concludes with Glock's 'Reflections and Replies' to the authors' contributions to the festschrift.

Acknowledgements

We would like to thank Hans-Johann Glock for taking an afternoon out of his busy schedule to give us an interview, which is the basis of the short biography provided here, and for his feedback on an earlier version of this introduction. We are also grateful to Paul Klur and Christoph Wagner for their help with editorial issues, as well as to Sam Schuman and Andrew Weckenmann from Routledge for their support.

References

Fuchs, W.R. (1972). *Denkspiele vom Reißbrett: eine Einführung in die moderne Philosophie*. Stuttgart: Deutscher Bücherbund.

Glock, H.-J. (1996). *A Wittgenstein Dictionary*. Oxford: Blackwell.

Glock, H.-J. (2000). Animals, Thoughts and Concepts. *Synthese* 123: 35–104.

Glock, H.-J. (2006). Concepts: Representations or Abilities? In Di Nucci, E. and McHugh, C. (eds.), *Content, Consciousness, and Perception: Essays in Contemporary Philosophy of Mind*. Cambridge: Cambridge Scholars Press, pp. 37–61.

Glock, H.-J. (2009). Can Animals Act for Reasons? *Inquiry* 52: 232–254.

Glock, H.-J. (2010a). Can Animals Judge? *Dialectica* 64: 11–33.

Glock, H.-J. (2010b). Concepts: Between the Subjective and the Objective. In Cottingham, J. and Hacker, P.M.S. (eds.), *Mind, Method and Morality: Essays in Honour of Anthony Kenny*. Oxford: Oxford University Press, pp. 306–329.

Glock, H.-J. (2010c). What Are Concepts? *Conceptus* 96, special issue 1: Predication and the Unity of the Proposition: 7–39.

Glock, H.-J. (2012). Letter from Switzerland. *The Philosopher's Magazine* 1st Quarter: 47–50.

Glock, H.-J. (2013). Animal Minds: Philosophical and Scientific Aspects. In Racine, T.P. and Slaney, K.L. (eds.), *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*. Basingstoke: Palgrave, pp. 130–152.

Glock, H.-J. (2014). Reasons for Action: Wittgensteinian and Davidsonian Perspectives in Historical and Meta-Philosophical Context. *Nordic Wittgenstein Review* 3: 7–45.

Glock, H.-J. (2017). Impure Conceptual Analysis. In Overgaard, S. and D'Oro, D. (eds.), *The Cambridge Companion to Philosophical Methodology*. Cambridge: Cambridge University Press, pp. 83–107.

Glock, H.-J. (2019). Agency, Intelligence and Reasons in Animals. *Philosophy* 94: 645–671.

Glock, H.-J. (2021). Concepts and Experience: A Non-Representationalist Approach. In Demmerling, C. and Schröder, D. (eds.), *Concepts in Thought, Action, and Emotion*. Abingdon: Routledge, pp. 21–41.

Glock, H.-J. and Schmidt, E. (2021). Pluralism about Practical Reasons and Reason Explanations. *Philosophical Explorations* 24: 119–136. https://doi.org/10.1080/13869795.2021.1908578.

Künne, W. (2003). *Conceptions of Truth*. Oxford: Clarendon Press.

Strawson, P.F. (1992): *Analysis and Metaphysics*. Oxford: Oxford University Press.

Part I

Wittgenstein

1 Farewell to hinge propositions

Severin Schroeder

In 1972, Georg Henrik von Wright published an article presenting the key ideas from Wittgenstein's final notebooks posthumously published under the title *On Certainty*. There Wittgenstein considers some truisms or certainties put forward by G.E. Moore in his attempt to refute scepticism, such as 'I have two hands', 'This is a hand', or 'The earth has existed a long time before I was born'. According to von Wright, Wittgenstein makes a number of important claims about such certainties. They are said to be:

i not known;
ii not empirical;
iii but rules of grammar;
iv ineffable;
v not propositional;
vi not truth-apt.

These claims have been developed and defended in greater detail by other commentators, in particular by Danièle Moyal-Sharrock (2007, 2017). Picking up one of Wittgenstein's metaphors for their transcendental role (OC § 341), such certainties have often been called 'hinge propositions'. The doctrine of hinge propositions has been regarded as a new development in Wittgenstein's thinking, so significant that the author of *On Certainty* has been called the 'third Wittgenstein' (Moyal-Sharrock 2004).

In his 2004 paper 'Knowledge, Certainty and Scepticism: In Moore's Defence', and again in his 'Philosophy Rehinged?' (2016), Hans-Johann Glock takes issue with that account of Moore's certainties, defending Moore's view that they are empirical and can be said to be known. I shall side with Glock in arguing that the idea of hinge propositions is, indeed, not the anti-sceptical panacea as which it has been presented. In spite of their certainty, Moore's truisms are fallible empirical propositions. To regard them as the basis of our language game is at best a metonymy, if not a category mistake. What is fundamental to our language game are no specific propositions (such as 'Here is a hand' or 'The earth

DOI: 10.4324/9781003202929-3

has existed for many years before my birth'), but the standards of rationality that make us regard such propositions as certain, at least under normal circumstances.

A note of exegetical caution

Von Wright's influential claims about 'hinge propositions' are not without some textual support. However, for each claim, there are also passages supporting the opposite view. That is because, as Glock notes (2004, 64)—unlike the *Tractatus* and *Philosophical Investigations*—*On Certainty* is not a book; it is simply a collection of (parts of) some first-draft notebook manuscripts (MSS 172, 174–177). They show Wittgenstein at work—thinking and jotting down his thoughts as they occur to him. They are exploratory and tentative, containing some raw ideas, rough approximations, semi-retractions, and contradictions. Since he never revised any of the material—never removed an initial claim that *p* in the light of his later observation that, actually, ~*p*—it would obviously be careless to ascribe to him as a considered position what is only tentatively expressed in a spontaneous remark, often prefaced by words such as 'I should like to say' (OC §§ 69, 75, 124, 151, 188, 308, 401, 404, 407, 414, 437, 447, 450, 457, 509) or followed by a self-critical rider such as 'And of course that is wrongly put' (OC § 303), 'But I am suspicious even of this. The sentence is too general' (OC § 321), 'That is very badly expressed and probably badly thought as well' (OC § 358), or 'But of course there is still a mistake even here' (OC § 405; cf. §§ 400, 402, 423, 437, 659, note after § 470). It is also worth taking into account what Wittgenstein says about his *On Certainty* remarks in general:

> I believe it might interest a philosopher, one who can think himself, to read my notes. For even if I have hit the mark only rarely, he would recognize what targets I had been ceaselessly aiming at.
>
> (OC after § 387; cf. § 532)

Clearly, Wittgenstein did not believe himself to have produced a new philosophical doctrine of basic certainties or hinge propositions.

You know you have hands

G.E. Moore tried to refute scepticism about the external world by giving a counterexample: an uncontroversial instance of knowledge of physical objects. He held up a hand and declared: 'I know that this is a hand.' So there is knowledge of a physical object after all. So, scepticism about the existence of the external world is false (Moore 1939).

Wittgenstein objected to Moore's procedure. Was that because he did not regard certainties such as the one that one has hands as *knowledge*?

'Because he adheres to the standard view of knowledge as justified true belief', yet my belief that I have hands is too basic for justification (Moyal-Sharrock 2017, 548; Coliva 2010, 142–143; Pritchard 2017, 573–574)? No, that is a misunderstanding.

Wittgenstein's concern was not to deny that Moore *knew* that he had a hand, but to point out that for Moore to *say* so was not enough—was simply begging the question. To refute someone takes more than contradicting them; to refute the sceptic requires more than a confident assertion that one does know things (OC §§ 389–390, 521). Wittgenstein explains:

> The wrong use made by Moore of the proposition 'I know …' lies in his regarding it as an avowal [*Äußerung*] as little subject to doubt as 'I am in pain'.
>
> (OC § 178; translation changed)

The grammatical peculiarity of avowals, such as 'I'm in pain', is that logically there is no room for error (PI § 288); truth coincides with truthfulness (PI II, 222). An expression of subjective certainty is an avowal, but a knowledge claim is not (OC § 389). Wittgenstein criticises Moore for conflating the two: for mistaking knowledge for a degree of certainty, for a mental state I can express with first-person authority (OC § 308; cf. Malcolm 1984, 71). In fact, I can be ever so certain and yet fail to know.

So, according to Wittgenstein, does Moore know that he has hands?—At one point, Wittgenstein registers an *inclination* to dispute that (OC § 407), but he immediately bethinks himself: Of course, one can know such things, even appropriately claim such knowledge. The following remark makes clear, again, that what really bothered Wittgenstein about Moore's assertion was that Moore seemed to take it as a refutation of scepticism (OC § 408), apparently taking the words 'I know' to express a mental state providing a guarantee of truth (cf. OC §§ 12, 14–15, 21, 404, 425).

What I know shows itself 'in the way I act and in the way I speak about the things in question' (OC § 395). And Moore's everyday behaviour would certainly manifest his knowledge that he has hands. What Wittgenstein is less confident about is whether Moore is correct to *say*: 'I know that I have hands': whether that is an *appropriate utterance in a philosophical discussion*.

> 'I know that this room is on the second floor, that behind the door a short landing leads to the stairs, and so on.' One could imagine cases where I should come out with this, but they would be extremely rare. But on the other hand I shew this knowledge day in, day out by my actions and also in what I say.
>
> (OC § 431)

Similarly, one can imagine circumstances in which Moore's trivial knowledge claims would sound natural enough (OC § 622), though without a suitable context their utterance would appear odd and unclear—just as my telling someone out of the blue that fish can swim. And yet, just as it is true that fish can swim, it is true that Moore has the knowledge in question (e.g., that he has hands): It shows easily in his everyday behaviour (e.g., his buying gloves, or his promising and planning to do things that require the use of his hands) (OC § 395).

What is Wittgenstein's account of knowledge? In *Philosophical Investigations*, we encounter the distinctive claim that knowledge presupposes the possibility of ignorance or error. Hence, it would be vacuous for me to say that I know that I am in pain (PI § 246; OC § 504). Pain and the awareness of pain cannot come apart, whereas a fact and the apparent knowledge of it must be logically independent. However, this requirement does not rule out Moore's knowledge claims. It is possible to have a hand without knowing it or to think one has a hand when one doesn't; so, the claim that I am neither unaware of my hand nor wrong in perceiving it is not logically vacuous. Of course, discrepancies between our possession of limbs and our awareness of them are extremely rare, so that we rarely have an occasion to make such a statement. But then, the contingent pointlessness of a statement—that it is just stating the obvious—does not seem enough to rule it out as nonsensical (cf. Glock 2016, 282). And in any case, we would hardly describe it as a misuse of a particular word. After all, an utterance of 'I know I have a hand' is not more otiose than other statements of the obvious, such as 'People have hands' or 'Pigs can't fly'. Far from being misuses of particular words, such truisms might even be used as examples in explanations of the meaning of words.

In his notebook discussions of 1950–1951, Wittgenstein considers some further analysis of the concept of knowledge. At one point, he suggests the following:

> One says 'I know' when one is ready to give compelling grounds. 'I know' relates to a possibility of demonstrating the truth. Whether someone knows something can come to light, assuming that he is convinced of it.
>
> But if what he believes is of such a kind that the grounds that he can give are no surer than his assertion, then he cannot say that he knows what he believes.
>
> (OC § 243; cf. § 91)

Such a requirement for knowledge would indeed appear to disqualify Moore's and many other elementary certainties. What could be more certain (and thus provide an effective epistemic ground for the assertion) than that I have a hand in front of me or that the colour of a ripe tomato is called 'red'? But Wittgenstein soon realised that such a verificationist

definition of knowledge is incorrect: It is not in accordance with out actual use of the word 'know'. For one thing, he notes that we can be said to know things for which we have no particular 'compelling grounds', but which are part of a world view that we have learnt about and that we take for granted:

> I know, not just that the earth existed long before my birth, but also that it is a large body, that this has been established, that I and the rest of mankind have forbears, that there are books about all this, that such books don't lie, etc. etc. etc. [. . .] This body of knowledge has been handed on to me and I have no grounds for doubting it, but, on the contrary, all sorts of confirmation.
>
> (OC § 288)

Those truisms are confirmed by lots of things we see and hear, but we could hardly cite anything in particular as a 'compelling ground' and more certain than those truisms themselves.

For another thing, a few months later Wittgenstein observes that memory affords us certain knowledge for which we cannot give further evidence:

> 'I know that for the last month I have had a bath every day.' [. . .] I *know* that I bathed each day and I do not derive that from some other immediate datum.
>
> (OC § 417)

In such a case, past experience is the cause of our knowledge, but it does not give us any citable evidence (cf. OC § 429). What is more, knowledge can also be based on experience that one cannot specify:

> One says 'I know that he is in pain' although one can produce no convincing grounds for this.
>
> (OC § 563)

And yet Wittgenstein emphasises that this is a case of *knowledge*, not just of subjective certainty: for it is a matter of 'understanding' (OC § 563).

Basic semantic knowledge is a particularly clear case of knowledge for which we cannot cite 'compelling grounds'. I know that in German the colour of ripe tomatoes is called '*rot*' (OC § 528), although no evidence I could provide (e.g., a dictionary entry) would do any work. After all, it is much easier to imagine a false dictionary entry than that a competent native speaker be mistaken about elementary colour words (cf. OC § 150).

In March and April 1951 (the last third of the remarks in *On Certainty*), Wittgenstein begins to put forward different (partial) analyses of knowledge, contradicting his earlier attempt of § 243 (before 23 September

1950). He observes that the word 'know' is often used to indicate that one is already in possession of a certain piece of information:

> I say 'I know p' either to assure people that I, too, am aware of the truth of p [. . .].
>
> (OC § 424; translation changed)

> 'I know that' may mean: I am already informed of it [. . .].
>
> (OC § 582; translation changed)

This is not a matter of the trustworthiness of my sources or whether I have conclusive evidence; what the knowledge claim expresses is simply that the news has reached me, that there is no need to tell me (cf. Hanfling 2000, 97). Another use of the word 'know' is to indicate reliability:

> [. . .] the purpose of the phrase 'I know' might be to indicate where I can be relied on.
>
> (OC § 575)

> 'I know what kind of tree that is.—It is a chestnut.' [. . .] The [. . .] 'I know' means roughly: I can say.
>
> (OC § 591)

Again, what makes it correct for me to claim knowledge is that I am in a position to give reliable information. There is no requirement that I also be able to *prove* the truth of my information. In accordance with that usage, as people do not need to be told that they have two hands and can reliably say so, it is correct to say that (normally) people *know* that they have two hands (OC § 430).

Empirical propositions

On von Wright's (and Moyal-Sharrock's) reading of *On Certainty*, Moore's truisms, or 'hinge propositions', only *appear* to be empirical propositions, but are, in fact, rules of grammar (von Wright 1972, 173; Moyal-Sharrock 2017, 553–555). Indeed, Wittgenstein suggests such a view:

> I am inclined to believe that not everything that has the form of an empirical proposition *is* one.
>
> (OC § 308; cf. § 213)

> I want to say: propositions of the form of empirical propositions, and not only propositions of logic, form the foundation of all operating with thoughts (with language).
>
> (OC § 401)

But we should note the cautious introductory words warning us that this is but a first approximation, not a considered position. Note also that § 401 is immediately followed by a partial retraction:

> In this remark the expression 'propositions of the form of empirical propositions' is itself thoroughly bad.
>
> (OC § 402)

And on the other hand, there are several remarks that seem to accept Moore's truisms as empirical propositions, e.g.:

> The *truth* of certain empirical propositions belongs to our frame of reference.
>
> (OC § 83)

> There are countless general empirical propositions that count as certain for us.
>
> (OC § 273)

And why should Moore's truisms—say,

(1) The earth has existed for many years before my birth.
(2) Here is a hand.

—*not* be regarded as empirical propositions? For one thing, their negation makes sense: It may not be credible, but it is easily imaginable (cf. Glock 2004, 72; 2016, 285). For another thing, they are based on experience. I *see* my hand, I *feel* it, and I have regularly done so for decades; and only because of those countless perceptions of my hand all my life do I not need to take a specific look now, but can confidently say that I have a hand even with my eyes closed. Even where such a truism is not *derived* from any particular observation, 'previous experience' would still 'be the cause of my present certitude' (OC § 429). Indeed, it is a very important point in Wittgenstein's philosophy of mind that we should not expect empirical propositions always to be derived from observational premises. Counterexamples he discussed are 'I am in pain', 'My left arm is bent behind my back', 'The noise came from the left', and indeed simple observational statements, such as 'This is a red patch' (cf. Glock 2004, 71; 2016, 284).

As for (1) 'The earth has existed for many years before my birth', it might be suggested that it cannot be *supported* by any historical evidence, since any historical evidence must *presuppose* it, i.e. the existence of a historical past (cf. OC § 411; Morawetz 1978, 18–19; Glock 2004, 72). A first reply is that support and presupposition are not mutually exclusive. Relying on a witness's account of what happened at a party

presupposes that he was there; but at the same time, his ability to give a detailed and plausible account of what happened is also evidence that he *was* there. It is true, however, that some specific pieces of historical evidence presuppose more general assumptions without supporting them. For instance, the minutes of a Cabinet meeting prove who was present—provided we take the document as minutes, as an accurate report, rather than as a piece of fiction. And of course, for that we need to know more than what it says in the document. But then, support for such more general assumptions, all the way up to the very existence of a historical past, is also available. *Psychologically* there may be a presupposition here—we take the past for granted, but *logically* there isn't (cf. Glock 2016, 276–277). We can ask: What in the light of our experience (of what we have actually seen of the origin, development, and duration of people and artifacts) is the more likely origin of, say, London and its inhabitants: (a) that people had parents who had parents themselves and so forth, and that buildings and other artefacts were developed and produced at roughly the speed at which we have seen such things being produced, or (b) that it all popped into existence shortly before my birth? Clearly, (a) is overwhelmingly likely in the light of our experience, whereas (b) 'the opposite hypothesis has *nothing* on its side' (OC § 190).

Rules of grammar?

At one point, Wittgenstein suggests something about certain apparently empirical propositions that he had elsewhere suggested about mathematical equations (RFM 324–325),[1] namely that they have been 'hardened into rules':

> It might be imagined that some propositions, of the form of empirical propositions, were hardened and functioned as channels for such empirical propositions as were not hardened but fluid.
>
> (OC § 96)

> Isn't what I'm saying: any empirical proposition can be transformed into a postulate—and then becomes a norm of representation.
>
> (OC § 321)

But again, the remarks are framed by notes of caution. 'It might be imagined' warns us that what follows is not actually Wittgenstein's conviction. And the quoted first sentence from § 321 is immediately followed by a retraction:

> [. . .] But I'm suspicious even of this. The sentence is too general [. . .]. It sounds all too reminiscent of the *Tractatus*.
>
> (OC § 321)

More importantly, it is not difficult to find reasons *not* to regard Moorean truisms as rules of grammar. First, a rule of grammar would be used for teaching or correcting the use of language, or to derive consequences from empirical propositions. Wittgenstein imagines a didactic use of (1) 'The earth has existed for many years before my birth'?—but, as he says, it would be comparable to teaching a child that a certain mountain is taller than a house, i.e. communicating some basic *empirical* knowledge (OC § 233). And if someone were to speculate, like Russell, that the world might have been created only shortly before his birth—a proposition like (1) would *contradict* such sceptical speculation, but it would not resolve it. Why should it be regarded as a rule, rather than a widely accepted empirical belief? And why a rule of *grammar*? Rules of grammar must be accepted as expressions of a consensus about the correct use of words. What words would be misused in Russell's sceptical question?

Secondly, (2) 'Here is a hand' may, indeed, be used to give an ostensive explanation of the meaning of the word 'hand': an example of a correct application of the word 'hand'. But to give an *example* of correct usage, a correct application of a predicate, is not the same as giving a *rule*, which would be an abstract general formulation of how to use a word (see Schroeder 2017, 254). 'A square is a plane figure with four equal sides and four right angles' is a rule (or definition); 'This is a square ☞ □' is not, it is just an exemplification. Large dictionaries contain both: They first define a word (rule) and then add examples of correct usage. It is important to note that any correct (empirical) application of a word can also be cited or referred to as a linguistic example for the benefit of language students. Every single sentence in *David Copperfield* could be cited in a dictionary to illustrate the correct use of a word or grammatical construction. That does not mean that Charles Dickens's novel consists only of rules.

Another justification for the view that Moore's truisms are not to be regarded as empirical propositions, but as rules, is that they are said to be fundamental to the language game: No doubt about them can exist 'if making judgements is to be possible at all' (OC § 308).

> 'I cannot doubt this proposition without giving up all judgement.'
> (OC § 494)

But is that true? Would reckoning with the falsity of such a 'hinge proposition' really undermine my capacity for judgement?

No, it is not difficult to imagine a situation in which Moore was hallucinating a hand that, in fact, had just been amputated, or indeed to imagine a very recent creation of the world with all its evidence of a long past. And it is also possible for somebody to develop doubts in either case, for example by being shown examples of sophisticated holograms

of missing limbs or by being exposed to several well-crafted science fiction films (or even pseudo documentaries) about the artificial creation of seemingly ancient worlds. Such doubts may not be altogether rational: They may evince an impressionable and credulous mind, but they do not amount to a sweeping collapse of all one's judgements. Such doubts may well be compartmentalised in an otherwise reasonable person, like somebody's whimsical obsession with the idea of aliens abducting aircraft from the Bermuda triangle.

In April 1951 Wittgenstein himself realised that his earlier sweeping claims about the fundamental role and indubitability of such truisms had been exaggerated and needed qualification:

> But since a language-game is something that consists in the recurrent procedures of the game in time, it seems impossible to say in any *individual* case that such-and-such must be beyond doubt if there is to be a language-game—though it is right enough to say that *as a rule* some empirical judgements or other must be beyond doubt.
>
> (OC § 519)

> When we say 'Certain propositions must be excluded from doubt', it sounds as if I ought to put these propositions—for example, that I am called L. W.—into a logic-book. For if it belongs to the description of a language-game, it belongs to logic. But that I am called L. W. does not belong to any such description. The language-game that operates with people's names can certainly exist even if I am mistaken about my name—but it does presuppose that it is nonsensical to say that the majority of people are mistaken about their names.
>
> (OC § 628)

It is not so difficult to imagine someone forgetting their own name, or getting seriously confused or doubtful about their name, while maintaining their linguistic competence and sound judgement in most other areas. And likewise for any other Moorean truism. No such *individual* proposition is a 'hinge' on which many other judgements turn. Against the sceptic, Wittgenstein maintains that global doubt makes no sense. Doubt requires *reasons* to doubt (OC §§ 4, 122, 323, 458; cf. Glock 2004, 73). And a *framework of certainty* is required for us to be able to raise a doubt about something (OC § 519). 'Some empirical judgements or other' [*irgendwelche Erfahrungsurteile*], and probably quite a few, must, at any time, be a reliable backdrop for our dealings with the world. However, no *particular* empirical proposition (of the kind presented by Moore) must always be part of that framework: absolutely immune from doubt. No such particular proposition has anything like the status of a rule (OC §§ 655–656).

Ineffable certainties?

Von Wright also suggests that, according to Wittgenstein, Moorean certainties are ineffable: an attempt to say what cannot be said (von Wright 1972, 174–176). He does not give any textual evidence from *On Certainty*, but he supports this claim by invoking the principle of bipolarity that Wittgenstein formulated in 1913, making sense of a proposition dependent on its contingency: its being able to be true or false (rather than necessarily true or necessarily false). Then, if we were to agree that Moore's truisms cannot be doubted and are therefore necessarily true, we might conclude that they lack sense, i.e. they are not something that can be said.

First of all, it should be noted that on this reading of *On Certainty*, far from presenting a 'third Wittgenstein', it would appear to be a relapse into the first. The *Tractatus* distinction between empirical (contingent) and conceptual (necessary) truths is, undoubtedly, a major insight. However, the further claim that only empirical propositions are meaningful is a youthful exuberance—engendering the delightful paradox that the *Tractatus* itself had to be dismissed as 'nonsense' (and keeping a generation of *Tractatus* scholars harmlessly occupied). Of course, conceptual truths have no empirical content. From an empirical point of view, they are vacuous. But that is misdescribed by the word 'ineffable': A lack of content does not mean that there is something unsayable there, it simply means that nothing is being said. And even that required the qualification: from an *empirical* point of view; for saying, for example, that a certain claim logically entails another is not saying nothing. The young Wittgenstein's conclusion that he did not really say anything in the *Tractatus* is plainly ridiculous, as brought out nicely by Russell's dry comment: 'Mr Wittgenstein manages to say a good deal about what cannot be said' (1920, 284).

The principle of bipolarity was rightly dropped in Wittgenstein's later philosophy when he realised that meaning is use and that necessary truths can have a use too—not the least in philosophy. Hence, even if (contrary to what I argued above) we did regard Moore's indubitable observations as necessary truths—that would not provide a good reason to consider them as devoid of meaning, let alone as intimations of the ineffable.

Yet Moyal-Sharrock continues to defend the ineffability claim, citing some passages in which Wittgenstein appears to deny that Moore's utterances make sense, e.g.:

> Thus it seems to me that I have known something the whole time, and yet there is no point [*habe es keinen Sinn*] in saying so, in uttering this truth.
>
> (OC § 466; translation changed)

Here, we should distinguish two things, which Wittgenstein occasionally runs together, namely: linguistic (utterance) meaning and the point or usefulness of an action (which may be a linguistic utterance) (cf. Glock 2016, 283). The standard expression in German for deficiency in the latter is the one used in § 466: '*es hat keinen Sinn*' (literally: 'it has no sense'). Thus, 'it has no sense' to repair a broken fence with a glue stick, to have a stern word with your cat for killing birds, or to tell somebody what you know that they already know. Those are pointless actions, a waste of time—but does it follow that they cannot be done? Of course not. By the same token, it would be rather odd to say of a verbal utterance—of saying something—that if it is pointless, then it cannot be said: is ineffable or unsayable. No, you can say it all right, it is easily sayable—only there is no point.

Does it follow from the pointlessness of a verbal utterance that it is linguistically meaningless? Are your indignant words to your predatory cat *linguistically* deficient? Is a widow addressing her husband's grave talking nonsense? Obviously not.—Sometimes, Wittgenstein was inclined to restrict the *use* that makes for linguistic meaning to *usefulness*; for example, in PI § 260, when he suggests that a note must have a 'function'. That is exaggerated. Although *typically* notes have functions, it is not uncommon for people to derive satisfaction from labelling or recording things for the sake of it (just think of train spotting). But usually, Wittgenstein was well aware that not every understandable human utterance is useful or functional. Often, we express thoughts and feelings without any communicative or instrumental purpose. Thus, it is perfectly normal for people to spontaneously say what they see to somebody standing next to them and seeing it just as clearly, for instance: 'It's raining' or 'There's the bus'. Again, think about soliloquies—speaking when there is nobody present to whom we could give any information. It is, in fact, very common for people to speak to themselves, sometimes just describing or labelling what they see in front of themselves. You look up and say to yourself: 'A squirrel'. What purpose does that utterance have? None. But such idle remarks are not that different from Moore's observation 'Here's a hand'.

Moreover, there appears to be an inconsistency in Moyal-Sharrock's assertion that hinge propositions are both 'ineffable' and 'rules of grammar' (2017, 549). The early Wittgenstein's claim that conceptual truths cannot be meaningfully said was dropped exactly because he realised that they do have a perfectly respectable use as rules of grammar. The idea of a rule of grammar discredited the bipolarity principle as a criterion of linguistic meaning, and hence the idea of ineffable necessary truth.

There are cases of intuitive understanding where the basis of our judgement cannot be specified. For instance, some people are reliable judges of character without being able to give us rules as to *how* to judge

a character (cf. LW II 61). Here, one may perhaps speak of something ineffable. It is the very opposite of the case where we can give a rule.

Finally, even if (contrary to what I argued above) we were to insist that in order to have meaning an utterance must have a communicative function or sense, it is not clear whether Moore's utterances are as incomprehensible as Wittgenstein suggests (e.g., OC § 347). Moore's truisms are not exactly rules, but they do play an understandable role in Moore's discussion of scepticism. According to Wittgenstein, in philosophy we 'assemble reminders' of trivialities in order to break the spell of philosophical doctrines or pictures (PI § 127). Trivial examples of pieces of knowledge are a case in point. When facing the sceptical view that one does not know anything, it is very natural to remind oneself of some straightforward examples of things we feel confident to claim we do know. As Wittgenstein rightly objected, that may not be an effective refutation of scepticism, but it is at least understandable as a somewhat naïve attempt at such a refutation. Thus, we do understand the communicative purpose of Moore's giving his humdrum examples.

Non-propositional certainties?

Related to the idea of ineffability is von Wright's further claim that Moore's certainties are 'not propositional knowledge' (1972, 178–179). Moyal-Sharrock calls this the most ground-breaking insight in *On Certainty*: 'basic beliefs being ways of acting (and not propositions striking us as true)' (2017, 556). However, the contrast is less than straightforward. There are, indeed, certainties manifest in our most primitive forms of behaviour (e.g., walking about, picking up things to eat) and preceding language use and ratiocination. Moyal-Sharrock speaks of 'animal trust in certain things' (2017, 559). But that a belief predates language does not mean that it cannot be expressed in language: by 'a proposition striking us as true'. Animals show awareness of their limbs, of obstacles in their way, of other animals, and of food to eat. Human beings can verbalise such basic certainties.

Moyal-Sharrock seems to deny that, asserting that such certainties 'can only manifest themselves as ways of acting'—but then immediately afterwards she states that they can be put 'into words for the benefit of philosophical elucidation' (2017, 557). Still, she insists that the fact that they can be verbalised does not 'make our certainties into propositions' (2017, 558). — Now, of course, somebody's certainty is not the same as a proposition, just as a desire is not a proposition. To identify a human attitude or feeling with a linguistic entity or abstraction (a meaningful declarative sentence or the meaning of a declarative sentence) would be a category mistake. In that sense (a), no certainty or belief is 'propositional', that is, identical with a proposition. However, on more

interesting readings, the question of whether a belief is propositional could mean:

(b) whether the believer is able to express their belief in words; or
(c) whether it is at all possible to express the belief in words; or
(d) whether the belief can only be ascribed to a language user.

On the second construal (b), a cat's awareness of another cat is non-propositional, whereas my awareness of that other cat is propositional. On (c), it is doubtful whether there can be any non-propositional beliefs, but there may be non-propositional anxieties: where a person cannot articulate what they are afraid of, but their behaviour shows a certain kind of fear or insecurity. On the final reading (d), the belief that Paris is the capital of France is propositional, just as the belief that water boils at 100°C, for they involve abstract concepts that can only be acquired with a system of symbolic representation. By contrast, a dog can be said to believe that there is someone at the door or that it is going to be given some food. These would be non-propositional beliefs. However, none of those construals fits the account of Moorean certainties as non-propositional. On (a), no belief whatsoever is propositional; on (b) and (c), Moore's certainties are clearly propositional; and on (d), some of them are—e.g. (1), as it involves abstract concepts of time measurement—some, like (2), are not.

The basis of our language game

And yet the idea that something non-propositional plays a key role in Wittgenstein's epistemological considerations is quite correct. Only it is not Moore's truisms or 'hinge propositions'.

As argued, truisms such as (1) and (2) are not as special as they have been cracked up to be. They are neither unknowable, nor rules of grammar, nor ineffable; they are empirical propositions all right (and as such clearly truth-apt). We do not doubt their truth, yet their falsity is not inconceivable (OC §§ 404, 663). And if one of them turned out to be false—if, for instance, the tree in front of me were only a hologram—I would neither go mad nor give up on all judgements. In fact, none of those trivial observational judgements (such as 'This is a tree') is aptly described as a hinge proposition; nothing much 'hinges' on such a judgement. What really plays a pivotal role in our epistemic relations to the world is not such a judgement, not any given proposition, but: *that we tend to make such judgements or tend to be certain of the truth of such propositions*. That is to say, what our language game 'hinges' on are not propositions, but the attitude that is reflected in our practice of judging—: that we regard some judgements as *reasonable*, and others as unreasonable (OC §§ 220, 254, 327, 334; cf. Glock 2016, 287).

And the attitude that informs a practice belongs to a different logical category from the results of actions that exemplify the practice. The musicality expressed in someone's practice of singing is not itself a song, and the rationality that informs our practice of judging in a certain way is not itself a judgement or proposition. Hence, categorially, it is quite true to say that the basis of our judgements is non-propositional; only it must not then be identified with Moore's propositions. Wittgenstein's remark about some propositions being 'like hinges' (OC § 341) is—like many remarks in those notebooks—only a first approximation, and taken at face value it is rather misleading.

I have already quoted the remarks where Wittgenstein effectively retracts the idea of hinge propositions (OC §§ 519, 628; cf. §§ 655–656). No such empirical proposition must be presupposed for our language game to function. Any one of them might be false without undermining the language game. What is a presupposition, however, is that by and large we are sufficiently reliable when making straightforward empirical judgements. A very similar idea has already occurred in *Philosophical Investigations*, at the end of the rule-following discussion:

> If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgements. This seems to abolish logic, but does not do so.—It is one thing to describe methods of measurement, and another to obtain and state results of measurement. But what we call 'measuring' is partly determined by a certain constancy in results of measurement. (PI § 242)

However explicit our rules and definitions, they can always be misunderstood. What really grounds our understanding is practice, ultimately based on examples (Z § 301; OC §§ 26–29) and sufficient uniformity in the way we tend to react to such teaching by examples and continue in our linguistic practice. The necessary stability of our linguistic practice cannot be guaranteed by rules alone; there must also be general agreement in their applications. Wittgenstein's famous example is that we all agree in our applications of the rule '+2', rather than some of us having strange ideas as to how to apply it to numbers above 1,000 (PI § 185). That we take the next number in the '+2' series to be 1,002 (and not 1,004) is comparable to our easily identifying a tree in front of us in broad daylight or a hand that is held up. The difference is that the mathematical example is just about the use of signs in a calculus, whereas in *On Certainty* the focus is on empirical judgements, which require not only a correct understanding of the meaning of the predicate involved, but also 'a certain constancy' in recognising the things it applies to. In the rule-following discussion, Wittgenstein was concerned with linguistic understanding (what does '+2' mean), whereas now the

focus is epistemological: not on whether you know what the word 'hand' means, but on whether you know that this really is a hand. Still, the fundamental idea from PI § 242 applies. If language is to be a means of communication, there must be regular agreement in judgements—reflecting both semantic agreement and factual agreement: about what, broadly speaking, the world is like. For instance, people must usually agree with Moore when he holds up a hand: both that it is called a 'hand' (not a 'foot') and also that it really is a hand and not just a hologram or a hallucination. In such shared attitudes, we manifest our standards of rationality. They are what is fundamental to our language game, not the specific propositions they make us express or accept on a given occasion (such as 'Here is a hand' or 'The earth has existed for many years before my birth').

Our standards of rationality could be expressed in something akin to grammatical propositions or norms of representation, which, however, are not those Moorean truisms themselves, but explanations of the circumstances that make them truisms. They are conditionals, of the form 'Under such circumstances … it is reasonable to assume that …' or 'This … counts as good evidence for …' (OC §§ 134, 334, 556). But then again, such conditionals cannot give a fully general expression of our standards of rationality, for either they refer to particular cases (as *examples* of what counts as good evidence, say: cf. OC § 124) or they contain expressions like 'under normal circumstances' which, again, we can only explain in terms of examples (OC § 674).

> Not only rules, but also examples are needed for establishing a practice. Our rules leave loop-holes open, and the practice has to speak for itself.
>
> We do not learn the practice of making empirical judgments by learning rules: we are taught judgments and their connexion with other judgments.
>
> (OC §§ 139–140)

Moore's and Wittgenstein's patently true judgements are not the foundation of our world view and language game, they merely exemplify it. The foundations are our common standards of rationality: of what kind of things a reasonable adult believes or accepts as true. Wittgenstein emphasises that those foundations are not true propositions—but an 'attitude' (OC § 404). In any particular case, however straightforward it appears, I may be wrong. Even about my hands or my name I am not infallible (OC § 425): One can always imagine an error due to some extraordinary circumstances. And yet, although I *could* be wrong, I am entitled to be certain that I cannot be wrong (OC §§ 663, 674). It is my *attitude* of certainty that matters and that is justified since without trusting some things no language game would be possible (OC § 509).[2]

Notes

1 For a critical discussion, see Schroeder (2021, 126–140).
2 I am grateful to Sam Couldrick, Christoph Pfisterer, John Preston, Robert Rabinowitz, Nicole Rathgeb, and Edoardo Sartore for comments on earlier versions of this paper.

References

Coliva, A. (2010). *Moore and Wittgenstein: Scepticism, Certainty and Common Sense*. Basingstoke: Palgrave Macmillan.

Glock, H.-J. (2004). Knowledge, Certainty and Scepticism: In Moore's Defence. In Moyal-Sharrock, D. (ed.), *The Third Wittgenstein*. Aldershot: Ashgate, pp. 63–78.

Glock, H.-J. (2016). Philosophy Rehinged? *International Journal for the Study of Skepticism* 6: 274–308.

Hanfling, O. (2000). *Philosophy and Ordinary Language: The Bent and Genius of Our Tongue*. London: Routledge.

Malcolm, N. (1984). *Ludwig Wittgenstein: A Memoir*. Oxford: Oxford University Press.

Moore, G.E. (1939/1959). Proof of an External World. In *Philosophical Papers*. London: George Unwin, pp. 127–150.

Morawetz, T. (1978). *Wittgenstein & Knowledge: The Importance of On Certainty*. Amherst: University of Massachusetts Press.

Moyal-Sharrock, D. (ed.) (2004). *The Third Wittgenstein: The Post-Investigations Works*. Aldershot: Ashgate.

Moyal-Sharrock, D. (2007). *Understanding Wittgenstein's on Certainty*. Basingstoke: Palgrave Macmillan.

Moyal-Sharrock, D. (2017). Wittgenstein on Knowledge and Certainty. In Glock, H.-J. and Hyman, J. (eds.), *A Companion to Wittgenstein*. Chichester: Wiley-Blackwell, pp. 547–562.

Pritchard, D. (2017). Wittgenstein on Skepticism. In Glock, H.-J. and Hyman, J. (eds.), *A Companion to Wittgenstein*. Chichester: Wiley-Blackwell, pp. 563–575.

Russell, B. (1920/1989). Introduction to *Tractatus Logico-Philosophicus*. Reprinted in: Wittgenstein, Kritische Edition. McGuinness, B. and Schulte, J. (eds.), *Logisch-philosophische Abhandlung*. Frankfurt am Main: Suhrkamp, pp. 258–286.

Schroeder, S. (2017). Grammar and Grammatical Statements. In Glock, H.-J. and Hyman, J. (eds.), *A Companion to Wittgenstein*. Chichester: Wiley-Blackwell, pp. 252–268.

Schroeder, S. (2021). *Wittgenstein on Mathematics*. London: Routledge.

von Wright, G.H. (1972/1982). Wittgenstein on Certainty. In *Wittgenstein*. Oxford: Blackwell, pp. 165–182.

Wittgenstein, Ludwig:

LW II *Last Writings on the Philosophy of Psychology*. Vol. 2. Edited by G.H. von Wright and H. Nyman. Translated by C.V. Luckhardt and M.A.E. Aue. Oxford: Blackwell, 1992.

MS Unpublished manuscript of Wittgenstein's Nachlass. Wittgenstein Archives at the University of Bergen. http://wab.uib.no/transform/wab.php?modus=opsjoner. Last accessed: 29 March 2022.

OC *On Certainty*. Edited by G.E.M. Anscombe and G.H. von Wright. Translated by D. Paul and G.E.M. Anscombe. Oxford: Blackwell, 1961.

PI *Philosophical Investigations*. 4th edition. Edited by P.M.S. Hacker and Joachim Schulte. Translated by G.E.M. Anscombe, P.M.S. Hacker and Joachim Schulte. Oxford: Wiley-Blackwell, 2009.

RFM *Remarks on the Foundations of Mathematics*. Edited by G.H. von Wright, R. Rhees, G.E.M. Anscombe. Translated by G.E.M. Anscombe, rev. ed. Oxford: Blackwell, 1978.

TLP *Tractatus Logico-Philosophicus*. Translated by D.F. Pears and B.F. McGuinness. London: Routledge & Kegan Paul, 1961.

Z *Zettel*. Edited by G.E.M. Anscombe and G.H. von Wright. Translated by G.E.M. Anscombe. Oxford: Blackwell, 1967.

2 Wittgenstein and Glock on history and historicism

Joachim Schulte

> One of my most important methods is to imagine a historical development of our ideas different from what has actually occurred. If we do that the problem shows us a quite new side.
>
> (Wittgenstein *Culture and Value*, 45)

Glock on historicism

As Hanjo Glock notes at the beginning of his essay 'Wittgenstein and History' (Glock 2005), there is very little to be found in the literature that might be regarded as a helpful discussion of this topic. At the same time, it is far from obvious that we may with confidence speak of anything like a 'topic' in this context. And this is so, not because we cannot think of any interesting connections between various ideas of history or historical views and notions, on the one hand, and Wittgenstein's philosophy, on the other, but rather because it is not clear as to which aspects of our standard questions about historical thought could be applied in a fruitful way to Wittgenstein's thought.

In view of this situation, it is particularly welcome that Glock has tried to address this admittedly vague subject-matter in a relatively systematic fashion. According to Glock, there is (as we have seen) hardly anything in the literature on the topic 'Wittgenstein and History'. Moreover, there is, as Glock says, very little in Wittgenstein's writings that might be considered relevant to the topic of history. As he puts it, 'there are a few scattered remarks. And there is also a certain amount of biographical evidence' (*ibid*., 236), where by 'biographical evidence' he probably means manuscript remarks by Wittgenstein that have not been published as parts of his, as it were, 'officially' philosophical writings or observations attributed to him in records of conversations with Wittgenstein.

As I shall try to explain in what follows, in my view there *is* a little more relevant material than just a few scattered remarks plus biographical evidence. But this material will become visible only if one takes a

DOI: 10.4324/9781003202929-4

somewhat different attitude from Glock's towards Wittgenstein and his writings. One main point of difference between Glock's perspective and my own way of looking at Wittgenstein (and his attitude to history, in particular) is that Glock tends to see Wittgenstein in the context of a school of 'analytic' philosophers, whereas I for my part (and in agreement with his own declarations and protestations) prefer to see him as belonging to no sort of school at all.[1] This, however, is simply a difference in point of view that may be usefully stated but does not need to be defended by arguments or evidence.

Most readers of Glock's article will have noticed the elegant way in which he adopts, or appropriates, the often perplexing term 'historicism'. He explains it in a way that rids it of any more substantial overtones that might, for example, remind us of theories involving a specific direction of historical development. Such theories would, for instance, suggest that human societies develop in cycles or that they advance from a primitive state to one of sophisticated perfection. Some of the best-known authors in the field of what, for want of a better word, one might call 'historical thought' have formulated and defended such theories, and, if I see it correctly, the expression 'historicism' is often used to refer to thinkers of this kind.

Another characteristic mark of a number of generally so-called 'historicist' thinkers is their tendency to make daring but insufficiently warranted predictions about future developments based on speculative assumptions about the course of history. Glock, however, decides to use this term in a less committal way and writes: 'For the purposes of this article I shall use the label "historicism" for any position which promotes historical thinking in philosophy and warns against ignoring or distorting the past' (Glock 2005, 238).

To be sure, this explanation has the advantage of sounding desirably unbiased and non-polemical, but it leaves readers a bit at a loss as to how to apply this term in a case like that of Wittgenstein. After all, what could it mean to say that Wittgenstein 'promotes, or fails to promote, historical thinking in philosophy'? There can be no doubt that Wittgenstein does say a number of things about philosophy and about recommended or banished ways of doing philosophy, but the status of these pronouncements is by no means obvious and hence requires a good deal of interpretation. In addition, even when a remark of this kind clearly involves mention of historical thinking, this has very little, or absolutely nothing, to do with history as discussed by Glock. An example would be an earlier version of a well-known remark on *Übersichtlichkeit* (PI § 122), where Wittgenstein points out that refraining from embracing a historical, fact-finding view and adopting an imaginative, fact-inventing way of proceeding[2] may be philosophically fruitful if we wish to gain a better understanding of certain internal relations (Wittgenstein PO, 133). But this implies no judgement about the usefulness of historical

thinking in philosophy—in the case at hand, this way of thinking is simply found to be beside the point.

And again, as regards the second part of Glock's explanation, it is difficult to see how this could fit in with what Wittgenstein does say and his general way of trying to make us see things. Why are these warnings attended to by Glock? Why should we be tempted to bring in questions about the past at all, and why should we be inclined to ignore or distort the past? Perhaps these questions sound a bit silly, but unless we have a vague idea about why one should raise and answer them, the 'historicism' understood in Glock's sense gets no grip. I suspect that such questions can appear meaningful when we are dealing, not with theories about the past and its ways of developing in general, but with certain philosophical theories of the past and our ways of representing and appreciating them. What goes to confirm this suspicion is the fact that Glock himself makes the apparently quite natural move of discussing certain philosophers' attitudes towards past philosophers and their theoretical achievements.

Spengler's influence

This aspect of the matter—that is to say, this decision to deal with the history of philosophy in preference to history simpliciter—would need further discussion. At this point, however, I want to express my (slight) puzzlement over the fact that Glock has found it prudent to give such a cautious explanation of the term 'historicism', even though there are passages in Wittgenstein's writings indicating that he was sympathetic toward stronger, or more substantial, forms of historicism. One obvious point is that in a well-known and often-quoted remark Wittgenstein states that Spengler (who, after all, was a kind of paradigmatic historicist in a strong sense of this word) was among those who influenced him (Wittgenstein CV, 16; 1931). And unless we find a successful way of attenuating the sense of 'influence', this statement surely involves an ambitious claim to the effect that Wittgenstein was sympathetic towards the reflections of a man who invented and promoted a forceful form of historicism.

But quite apart from this 'biographical' piece of evidence concerning Wittgenstein's attitude to historicism, there are several passages to be found in his manuscripts suggesting that he approved and accepted central theoretical claims and distinctions conceived by Spengler. The latter's distinction between *culture* and *civilisation* is a case in point. In Spengler's view, culture is associated with the flourishing of a society, civilisation (on the other hand) with its feeble ending. It seems to be precisely in Spengler's sense that Wittgenstein remarks:

> In the metropolitan civilisation the spirit can only huddle in some corner. And yet it is not for instance atavistic and superfluous but

> hovers above the ashes of culture as an eternal witness—as if an avenger of the deity. | As if it were awaiting a new incarnation (in a new culture).
>
> (Wittgenstein PPO, 55)

Perhaps unsurprisingly, there are several further passages from this period (ca. 1930) employing Spengler's distinction. But even as late as 1947, we find the following, characteristically Spenglerian, reflection:

> Perhaps one day a culture will arise out of this civilisation. | Then there will be a real history of the discoveries of the 18th, 19th and 20th centuries, which will be of profound interest.
>
> (Wittgenstein CV, 73; 21.12.1947)

It is difficult to make clear sense of this remark, but the beginning is intelligible enough: Wittgenstein regards it as possible that our civilisation—that is to say, this late and almost degenerate stage of the development of our society—will at some point of the future allow this society to turn into a culture. In other words, the envisaged possibility is that of a flourishing (as opposed to a static or decaying) society that is capable of bringing forth impressive achievements in fields of human activity that are apt to express the spirit of that culture. Of this culture, it is said that it will be able to engender a 'real history' (perhaps a faithful historical account) of human *Erfindungen* (literally: inventions—perhaps 'creations' or 'products of the imagination' are meant).

At any rate, what is striking about this passage is that, on the one hand, Wittgenstein makes use of the Spenglerian distinction between culture and civilisation but, on the other hand, does not buy into the theoretical background assumptions that are supposed to give that distinction its specifically Spenglerian flavour. In a way, Wittgenstein seems to ruin the whole idea by brushing aside its historical underpinnings that are supposed to be resting on the view that societies are bound to develop in accordance with a pattern of the type: youth, flourishing, maturity, and civilisation. And Spengler as well as his true followers claim that the possibility of a culture's growing out of a civilisation is simply not to be taken into consideration: A new culture will have to emerge from a new society, but not from a declining society that has long ago left behind its stage of cultural flourishing.

A great organisation and its spirit

It will be worth bearing in mind this unorthodox use that Wittgenstein makes of the distinction between culture and civilisation when we get down to considering the following extraordinary document written in November 1930, bearing the title *Zu einem Vorwort*, translated as

'Sketch for a Foreword' (Wittgenstein CV, 8–11). But what Wittgenstein means is possibly something like 'Materials for a Foreword' or 'Apropos of a preface'. It is from Peter Winch's translation of this document that I quote the following fragment:

> This book is written for those who are in sympathy with the spirit in which it is written. This spirit is, I believe, different from that of the prevailing European and American civilisation. The spirit of this civilisation the expression of which is the industry, architecture, music, of present day fascism and socialism, is a spirit that is alien and uncongenial to the author. This is not a value judgement. It is not as though I did not know that what today represents itself as architecture is not architecture and as though he [= the author] did not approach what is called modern music with the greatest mistrust (without understanding its language), but the disappearance of the arts does not justify a disparaging judgement on a whole segment of humanity. For in these times genuine and strong characters simply turn away from the field of the arts and towards other things and somehow the value of the individual finds expression. Not, to be sure, in the way it would at a time of Great Culture. Culture is like a great organisation which assigns to each of its members his place, at which he can work in the spirit of the whole [. . .] Even if it is clear to me that the disappearance of a culture does not signify the disappearance of human value but simply of certain means of expressing this value, still the fact remains that I contemplate the current of European civilisation without sympathy, without understanding its aims if any. So I am really writing for friends who are scattered throughout the corners of the globe.[3]
>
> (Wittgenstein CV, 8–9; 6.–7.11.1930)

Of course, it is true that Wittgenstein's words are often difficult to understand and require interpretation if we are to make clear sense of them. But some parts of this passage are particularly hard to grasp, owing to the fact that, as readers of Wittgenstein's philosophical remarks, we are not prepared for this type of general reflection. To mention just one point, hardly any readers of Wittgenstein's works will expect him to speak in this uninhibited way about what he calls 'the spirit' in which his book is written, which prevails in our civilisation, and which is characteristic of typical expressions of our time. What does he mean by 'spirit'?

Wittgenstein's first sentence seems to identify 'spirit' with a certain attitude that a person can assume and express in producing something to which people can, in their turn, respond in all kinds of ways, for example by feeling sympathetic towards that attitude. And this reading of 'spirit' (as, roughly speaking, an *attitude*) chimes well with the fact that it is the author of the book in question, and hence an individual

human being, who is spoken of as the bearer of the spirit in which the book is written.[4]

In talking about a person's attitude, we tend to assume that we are able to describe situations said to be instances in which that attitude is being expressed. And we tend to assume this for the reason that there are characteristic patterns we can appeal to in describing such situations. Thus, on the one hand, there is the friendly face of a person who is sympathetic towards what he is observing; on the other hand, there is the wrathful face of a person who is angry about the object of his attention.

Accordingly, one might try to argue that, in his first sentence, Wittgenstein is speaking about the attitude of a particular human being, namely himself as the author of his envisaged book; however, in his second sentence, he uses the word 'spirit' in talking metaphorically about a specific present-day attitude typically finding expression in the civilisation whose products we can observe in America and Europe. But what about the third sentence? Is this an extension of the same metaphorical use, or is it yet another kind of more or less metaphorical use?

In trying to answer this question, it may be helpful to go back to Wittgenstein's second sentence to see whether we have been hasty in speaking of a metaphorical extension of the first use of the word 'spirit'. The spirit mentioned at the beginning of the second sentence seems to amount to what we have called Wittgenstein's *attitude* in writing his book; and in view of his continuing his observations by talking about a 'different' but somehow comparable spirit, we decided to suppose that his second use of the term 'spirit' is a metaphorical extension of the first: an extension that would allow us to speak intelligibly about the spirit of our civilisation. I hope that readers will by now have seen what makes this move so tempting: We start from the assumption that talking about the author's spirit is the as it were 'easy' case, the individual case, which can be described in objective terms taken from the field of our everyday descriptions of attitude expressions. Once we have made sure that we feel at home in this area of descriptions, we believe that there is no obstacle to continuing in the same vein and thus to transposing our way of speaking of the individual spirit to the spirit of our civilisation.

Attentive readers might have noticed some time ago what I am driving at: My suggestion is that we are not dealing with a case of metaphorical extension, and that the explanation goes in the opposite direction. It does not go from the individual to the general, or the abstract, case; it goes from the general, or the abstract, case to the individual one. In other words, the spirit of our civilisation comes first, as it were, and its expressions are products that we respond to in ways that are similar to, but by no means the same as, our ways of responding to individual people as well as to their utterances and artefacts.

If this way of reading Wittgenstein's remark is roughly right, we shall have to read his observations on the expressions of our civilisation as

direct, that is to say: unmediated and non-metaphorical statements about developments at *a certain place* and at *a certain time*. One consequence of this insight is that we have to find and, if possible, to explain a non-metaphorical use of the expression 'spirit'. As we have seen, Wittgenstein does not hesitate to talk about expressions of the spirit of our civilisation, and this may lead us to imagine that those expressions stand to the spirit in a straightforward sort of causal relation: that the music of our time, for instance, is the direct work of an entity called 'spirit' that manifests itself in the music written today as well as in the forms of industrial production prevailing in our time. But it suffices to allude to this notion to see that this view is far too naive to be fruitfully ascribed to Wittgenstein.

Gestalt terms

On the one hand, it is clear that we have to take Wittgenstein's way of talking about the expressions of the spirit of our civilisation seriously. On the other hand, we have to avoid saddling him with a view that is too naive and simply contrary to the 'spirit' of his whole way of thinking. So, let us have another look at the critical term 'expression': We speak of sighs as expressions of dejection or relief, of cries as expressions of pain or horror, of a smile as an expression of cheerfulness, etc. These expressions can be perceived directly by looking at people's faces and listening to their utterances. In this case, *A*'s expression of pain will be distinct from *B*'s expression, no matter how similar they happen to be.

But there is another way of talking about expressions, and this is a way of talking in terms of which expressions can count as the same even if their utterers are different people and the means of their production are different instruments. In this sense, one can, for example, say that the horror to be seen on Anthony Perkins' face is the same as that expressed on a given page of Kafka's book. And of course, there is a corresponding way in which expressions in this second sense can be distinct from other expressions in this same sense. And just to have a handy label to mark the relevant distinction, we may speak of expressions in the *Gestalt* sense when we are dealing with expressions in the second kind of sense and of expressions tout court when dealing with the other kind of expressions.

Now I should claim that in asserting, for example, that our form of industrial production is an expression of the spirit of our civilisation I am using the term 'expression' in what I have vaguely called the *Gestalt* sense of this word. One consequence of this move is that (as we have just seen) in this case the statements of similarity and dissimilarity, of sameness and distinctness that will count as true or false will be completely different ones from those that count as true or false in the case of using non-*Gestalt* terminology. Another consequence is the fact that context will play an entirely different and a surely more complicated role: Taken in the *Gestalt*

sense, in scene *Z*, the expression on Anthony Perkins' face can be completely different from the expression on his face in scene *Y*, even though the director has used the same celluloid material for both scenes. Taken in the non-*Gestalt* sense, we have nothing corresponding to this possibility.

But in what way can these considerations help us to gain a better understanding of Wittgenstein's notion of the 'spirit' whose expressions the products of industry, architecture, and so on are supposed to be? This may be the point at which the idea of history, or rather the idea of historical development, comes to play a decisive role. After all, it is the specific spirit of the present day—of what Wittgenstein calls the *Jetztzeit*—that in his view is alien and uncongenial. And he emphasises that this judgement is not a value judgement. In other words, his statement about this spirit's being alien and uncongenial to him is not merely a statement of his dislike for this particular period and its products. On the other hand, it is only by being the spirit of a specific time and cultural context that its products can elicit this response from him in his role as a critical observer of those products.

Here, it is important to see that Wittgenstein distinguishes between periods of 'Great Culture' and periods 'without culture' (*Unkultur*, so perhaps 'barbarism'). Culture, he says, is like a vast organisation that 'assigns to each of its members his place, at which he can work in the spirit of the whole', and his powers can be judged by finding out about the magnitude of his contribution to this whole. At this point, the text of Wittgenstein's remark poses a problem: In speaking about culture as a great organisation (or, perhaps, an organising principle), it is not quite clear whether he is referring to certain 'Great Cultures' in particular or cultures in general—that is to say, as including periods of *Unkultur*. The tone in which he continues his remark and talks about the 'fragmentation' of forces and the 'wasting' of our strength characteristic of the times of *Unkultur* suggests that what he has in mind is *Great Culture* as opposed to *any culture*. But in the last analysis, it does not matter greatly as to whether one reads Wittgenstein's remark in this way or that way: It can be made sense of as long as we are prepared to accept the idea that a historical period can, or should, be regarded as a complex structure consisting of its members' contributions to the achievements of this period.

The crucial move is made by conceiving of the organising principle as the 'spirit' of the relevant time: This principle is active during a certain period of time—it is neither timeless nor sempiternal. It is defined through its products and the way these products are connected, that is, the specific way they fit together and can be said to harmonise with each other. And again, we find ourselves in the realm of *Gestalt* concepts and come to see that the expressions of the spirit of a particular time are products of this spirit, not in the sense of their being causally brought about by a more or less mysterious entity called 'spirit', but in the sense of being informed, or suffused, by the spirit of its time in such a way that

we find it possible to tell each other in an intersubjectively intelligible manner what we mean by 'fit together', 'harmonise', etc.

Readers of Wittgenstein's remarks may be struck by the fact that in listing some forms of expression of the spirit of the present day he never mentions philosophy. What he does mention are industry, architecture, and music as well as fascism and socialism.[5] Of architecture (which was his own occupation until a year or two before writing the remark we are discussing here) he claims to *know* that what today 'poses as [or pretends to be] architecture is not architecture' at all, and of 'modern' [that is, contemporary] music he says, first, that he 'approaches it with the greatest mistrust' and, second, that he does not even understand 'its language', which may mean that he would be hard put to give a description of the impression he might receive from listening to this kind of music if he did listen to it. These and several further observations on contemporary activities and their products make it clear that in Wittgenstein's view the times he lives in cannot be seen as the times of a 'Great Culture'. His description of the present period shows not only his lack of sympathy with and understanding of the products of this period, but also that he feels completely out of tune with the aims of the people surrounding him. This is why he ends this part of his draft by stating that basically he is 'writing for friends who are scattered throughout the corners of the globe'—and this obviously means that he needs the idea of the existence of friends out there in order to be able to continue his work as a writer of philosophical remarks.

Architecture as a gesture

In the previous section, I have insisted that at least some of Wittgenstein's observations framed in terms such as 'spirit' and 'expressions of the spirit' can be made better sense of if the relevant words are understood as *Gestalt* terms. Of course, this is just an echo of Wittgenstein's own insistence when writing of architecture, for example, that 'it is a *gesture*' (Wittgenstein CV, 49; 28.10.1942). What is not so easy to account for is the continuation of these words. Wittgenstein writes: 'Not every purposive movement [*zweckmäßige Bewegung*] of the human body is a gesture. Just as little as every functional building [*zweckmäßige Gebäude*] is architecture' (*ibid.*).

Clearly, a building only merits the title of 'architecture' if it displays qualities that make it stand out from other buildings. These other buildings (office buildings, barns, etc.) may fulfil their various purposes more or less well, but they exhibit no features in virtue of which they might be judged as (good or bad) architecture. Similarly, there are movements made by human beings that, even though they may be identified as such in virtue of their specific points or purposes, will never be called 'gestures' on account of the fact that they lack the required solemn or celebratory character.[6]

But to be sure, this does not mean that what Wittgenstein is driving at is that, if a given building is to deserve the label 'architecture', it will have to fulfil certain conditions that may qualify it as architecture. Rather, what he must have in mind is that if a building is regarded as a work of architecture, it must be possible to characterise it in 'the language of gestures'—that is to say, in *Gestalt*-like terms. As Rhees reports, Wittgenstein described works of architecture by saying that they 'followed a tradition by taking over certain rhetorical forms, but it says nothing in them' (Rhees 2016, 863). On another occasion, he said about the Georgian architecture in Dublin: 'The people who built these houses had the good taste to know that they had nothing very important to say; and therefore they didn't attempt to express anything' (Drury 2016, 809)—and here, of course, the allusion to the (gesture of) refraining from attempting to express anything is intended to bring out a *visible* quality of those buildings. On yet another occasion, Wittgenstein, rather characteristically, observed that 'the houses on a modern by-pass [. . .] had such "a silly smile"' (Lee 2016, 484). But what these remarks (and many others) may help to bring out is that *Gestalt* terminology tends to go hand in hand with expressions meant to situate the architects in question in a specific historical context ('they followed a certain *tradition*', 'the people who built those *Georgian* houses', 'the houses on a *modern* by-pass', etc.).

In the remark from which we excerpted our long quotation, Wittgenstein compares a Great Culture with 'a great organisation which assigns to each of its members his place, at which he can work in the spirit of the whole' (Wittgenstein CV, 8–9; 6.–7.11.1930). A similar idea is used much later (in October 1948) in a context where Wittgenstein wonders about the fate of major artists working in periods unfavourable to their art. Here, however, he does not use the image of an organisation's assigning a *place* to its members but the very similar one of assigning a *task* to them. The example Wittgenstein has in mind is that of Eduard van der Nüll, who (among other things) built the Viennese *Hofoper* and was one of the best-known representatives of the architectural school of historicism—a school whose period Wittgenstein evidently regarded as a despicable or pathetical one, while he had a high opinion of van der Nüll's abilities as an architect. It is in this sense that he wrote:

> A great architect in a bad period (van der Nüll) has a quite different task from that of a great architect in a good period. You must again not let yourself be deceived by the generic term. Don't take comparability, but rather incomparability, as a matter of course.
>
> (Wittgenstein CV, 84; 19.10.1948)[7]

The similarities between the situation described here and the situation of Wittgenstein himself in relation to his own times as described in our

long quotation from 1930 are obvious. But what is 'the generic term' of which he claims that it is apt to 'deceive' us?

In the present context, 'the generic term' that Wittgenstein is alluding to must be 'architect' or, perhaps, 'great architect'. If we use such terms in their standard ways, we expect to be able to make comparisons between the items instantiating the relevant concepts. That is, if we make comparative judgements about the works of van der Nüll and those of Brunelleschi or those of Mies van der Rohe, for that matter, we expect these judgements to be legitimate: After all, what we are comparing are the works of supposedly great architects, and hence works that, in this sense at any rate, belong to the same category. But as Wittgenstein points out, there is no good reason to expect those works to be comparable just because we agree in calling them works by architects, or great architects. As the architects mentioned belong to entirely different periods, we have to allow for the possibility that their works are not comparable at all. The contexts in which they tried to achieve something may have been too different to permit anything like a useful comparison. One reason is this: that the meanings of the descriptive terms we employ in making comparisons can vary to an extraordinary extent depending on the historical context in which they are used. And perhaps they vary to such a degree that the comparisons made on this basis prove deceptive.

Symbolic value and its historical context

One of Wittgenstein's examples of period-relative terms is 'drinking', and about drinking he says that 'at one time it is symbolic, at another boozing' (Wittgenstein PPO, 33). Here, the suggestion is that drinking may be embedded in completely different functional contexts whose possible interpretations vary to such a degree that we may be inclined to speak of incommensurability ('incomparability'). Similarly, in the case of architecture, some of van der Nüll's solutions, for example, may superficially look like certain solutions pioneered by Brunelleschi, but owing to differences in their respective historical, and hence social, frameworks it is possible that the descriptive terms used to characterise those solutions have completely different connotations or go to show that what seems to be a 'solution' in one case does not deserve this name in another.

Once he has made his remark about drinking, Wittgenstein continues by saying that 'the nimbus, or the genuine nimbus, does not attach to the external fact, that is, not to the fact' (Wittgenstein PPO, 33). By 'nimbus', I take it, he means something like the symbolic value of the activity concerned. Thus, drinking may be part of a ritual or of a number of ritual actions, and the participants may, for instance, conceive of these actions as ways of worshipping the gods. Their identification of such actions with elements of the ritual in question may be extreme and lead them to deny that what they are doing is a form of drinking at all.

This is why Wittgenstein feels that he may claim that 'the nimbus' does not attach to the external fact (i.e., the fact of drinking) but, as we are perhaps allowed to continue, to the purely symbolic action of worshipping the gods in this way.

At this point, it is interesting to observe Wittgenstein moving from 'the external fact' to the mere 'fact'. By putting things this way, he possibly wants to underline that there is no independently identifiable and purely factual element to which a 'nimbus' or any other form of symbolic embroidery could be said to attach. In the area Wittgenstein is talking about, there are no such supposedly 'external facts' but only actions that are deeply suffused with symbolic meanings, which, taken altogether, belong to a certain historical context where they form a complex, *Gestalt*-like whole whose constituents acquire sense and symbolic function by standing in certain relations to each other.

Much of what we have said so far in giving an account of certain historical, if not historicist, aspects of Wittgenstein's reflections is clearly in harmony with the following no doubt historical and perhaps historicist remark, which was written in the same year as our previous long quotation:

> We view the Copernican discovery as something great—because we know that it signified something great in its time and perhaps because a resonance of this significance comes across to us—and now we infer by analogy that Einstein's discoveries etc. are something at least equally great. But they are—no matter how great their practical value, many-sided interest etc.—only as great as they are significant (symbolic). It is with this of course as it is—for example—with heroism. A feat of weaponry of former times is—rightfully—praised as a feat of heroism. But it is quite possible that an equally or even more difficult feat of weaponry is today purely a matter of sport and is unjustly called a feat of heroism. The difficulty, the practical significance, all that can be judged, as it were, from outside; the greatness of the heroism is determined by the significance of the action. By the pathos which is associated with the way of acting.
>
> Because, however, a particular period of time, a particular race[8] associates its pathos with very particular ways of acting, people are led astray and believe that the greatness, significance lies necessarily in that way of acting. And this belief is always reduced to absurdity just when a transvaluation of values [*Umwertung der Werte*] comes about through an upheaval, that is, when true pathos now settles upon another way of acting. Then—probably always—the old, now worthless bills remain in circulation for some time and people who are not quite honest pass them off as great and significant until one finds the new insight once again trivial and says 'of course these old bills are worthless'.
>
> (Wittgenstein PPO, 31–33)

By now, I think, it is fairly easy to discern the pattern behind Wittgenstein's way of reasoning: On the one hand, we have Copernicus, his discovery, and an audience schooled in grasping the wider significance of things; on the other hand, there are Einstein and his audience of people interested in the practical side of things. The first group of people corresponds to those who go in for heroic forms of fighting, whereas the second group corresponds to people who do their fighting as a kind of sport. These groups are very different in kind, and as we have seen, they can differ to such an extent that they become mutually incomparable. The first group lives in a world of symbols, as it were, whereas the second group has no, or insufficient, access to this world. Members of the latter group fail to see the nimbus attaching to certain ways of acting appreciated by members of the first group, who in their turn are quite incapable of seeing those actions as 'mere facts'. On the contrary, they approach these actions with a special kind of 'pathos', as Wittgenstein says, by which he evidently (and in agreement with the standard use of the German word '*Pathos*') means a solemn form of impassionedness or fervour, allowing those who are seized by it to intuit the true value of symbolic ways of acting. Times differ according to which of the two groups is dominant and shapes the spirit prevalent at the relevant time and place. But once the feeling for symbolic actions that seems to shape the spirit of 'good' periods begins to dwindle, all kinds of misunderstandings and deceit tend to spread and to make it ever more difficult to arrive at a true appreciation of past achievements.

To be sure, this is a very rough sketch after a vague picture drawn by Wittgenstein in some of his remarks. He does not say that he sees himself as belonging to one of those groups. But he does say, or almost say, that he feels out of tune with a world in which so little remains of former 'good' periods. And as regards the purposes of the present paper, it seems clear that the historicism I would want to attribute to Wittgenstein on the basis of that picture is of a much stronger, much more substantial kind than that he is credited with by Glock. At the same time, I do not think that any of the other variants of historicism distinguished by Glock would really fit the picture I have tried to outline. So I suppose, we shall have to settle for some kind of incommensurability between our pictures and perhaps wait for a third party to show us where we may have gone wrong.

Notes

1 See, for example, the following remark from 1947: 'I cannot found a school, because I actually want not to be imitated. In any case not by those who publish articles in philosophical journals' (Wittgenstein CV, 69; 13.–14.4.1947); cf. Rhees (2016, 866).
2 Cf. the motto quoted at the beginning of this chapter.
3 It is characteristic that Georg Henrik von Wright chose this last sentence as his motto (1982).

4 There is a good deal more on Wittgenstein's use of the word 'spirit' (*Geist*) in my (Schulte 2011), see § 9.
5 As far as I can see, these (Wittgenstein CV, 8; 6.–7.11.1930) are the only occurrences of the words 'fascism' and 'socialism' in Wittgenstein's *Nachlass*.
6 Cf. (Wittgenstein PO, 129): 'One could almost say that man is a ceremonial animal.'
7 Roughly speaking, the second half of the nineteenth century was the heyday of architectural historicism. Probably it is not an accident that this period is one that Wittgenstein explicitly wishes to *exclude* from his cultural ideal (Wittgenstein CV, 4; 10.10.1929). As Brian McGuinness (2018, 24) points out, this wish seems to be incompatible with Wittgenstein's great admiration for Brahms and Bruckner. But perhaps this fact should rather help us to see Brahms and Bruckner as 'great composers in a bad period'. This might contribute towards arriving at a more fair-minded judgement about these composers and their merits.
8 Perhaps Wittgenstein is using the word 'race' in a sense resembling that in which he speaks of '*eine Menschheit*' in our first long quotation (there, Winch translates this freely, and maybe correctly, as 'a whole segment of humanity').

References

Drury, M.O'C. (2016). Conversations with Wittgenstein. In Flowers, F.A. and Ground, I. (eds.), *Portraits of Wittgenstein*. London: Bloomsbury, pp. 776–837.

Glock, H.-J. (2005). Wittgenstein and History. In Pichler, A. and Säätelä, S. (eds.), *Wittgenstein: The Philosopher and His Works*. Bergen: Working Papers from the Wittgenstein Archives at the University of Bergen, No. 17, pp. 236–262.

Lee, D. (2016). Wittgenstein 1929–1931. In Flowers, F.A. and Ground, I. (eds.), *Portraits of Wittgenstein*. London: Bloomsbury, pp. 476–485.

McGuinness, B. (2018). Einleitung. In McGuinness, B. and Schweitzer, R. (eds.), *Wittgenstein: Eine Familie in Briefen*. Innsbruck: Haymon, pp. 7–37.

Rhees, R. (2016). Postscript. In Flowers, F.A. and Ground, I. (eds.), *Portraits of Wittgenstein*. London: Bloomsbury, pp. 838–867.

Schulte, J. (2011). On a Remark by Jukundus. In De Pellegrin, E. (ed.), *Interactive Wittgenstein*. Essays in Memory of Georg Henrik von Wright, pp. 183–208.

von Wright, G.H. (1982). Wittgenstein in Relation to His Times. In *Wittgenstein*. Oxford: Blackwell, pp. 201–216.

Wittgenstein, L. (1953/2009). *Philosophical Investigations* [=PI]. 4th revised edition. Edited and translated by G.E.M. Anscombe, P.M.S. Hacker, and J. Schulte. Chichester: Wiley.

Wittgenstein, L. (1993). Remarks on Frazer's *Golden Bough*. In Klagge, J.C. and Nordmann, A. (eds.), *Philosophical Occasions, 1912–1951* [=PO]. Indianapolis, IN and Cambridge: Hackett, pp. 115–155.

Wittgenstein, L. (1998). *Culture and Value* [=CV]. Revised edition. Edited by G.H. von Wright, H. Nyman, and A. Pichler. Translated by P. Winch. Oxford: Blackwell.

Wittgenstein, L. (2003). Movements of Thought: Diaries, 1930–1932, 1936–1937. In Klagge, J.C. and Nordmann, A. (eds.), *Public and Private Occasions* [=PPO]. Lanham, MD: Rowman & Littlefield, pp. 3–255.

3 Wittgenstein's later nonsense*

Daniel Whiting

Preamble

Hanjo Glock was the primary supervisor of my doctoral research at the University of Reading, which (to my surprise!) began nearly 20 years ago. He was an inspirational supervisor—excitable, energetic, and enthusiastic—and hugely supportive to the students under his supervision. His vast knowledge of analytic philosophy and its history was an invaluable resource to draw upon, while his irreverent (and, at times, unrestrained) wit offered welcome relief from the uncertainties of early academic life.

In our many meetings, Hanjo's characteristic response to a piece of work I had shared, and in which I had put forward some view or other, was to ask, in a tone as impatient as it was good-humoured, 'But where is the argument!?' It seems, then, a fitting tribute to present Hanjo with what I hope he will recognise as arguments, whether good or bad, against some views that he himself has put forward.

Introduction

Nonsense figures prominently throughout Wittgenstein's philosophy. In the earliest entry in the *Notebooks*, he writes:

> Let us remember the explanation why 'Socrates is Plato' is nonsense. That is, because *we* have not made an arbitrary specification, NOT because a sign is, shall we say, illegitimate in itself!
>
> (NB, 22 August 1914)

And in a remark written in the final years of his life—between 1949 and 1951—Wittgenstein asks:

> But is it an adequate answer to the scepticism of the idealist, or the assurances of the realist, to say that 'There are physical objects' is nonsense? For them, after all, it is not nonsense. It would, however, be an answer to say: this assertion, or its opposite, is a misfiring

DOI: 10.4324/9781003202929-5

> attempt to express what can't be expressed like that. And that it does misfire can be shewn; but that isn't the end of the matter. We need to realize that what presents itself to us as the first expression of a difficulty, or of its solution, may as yet not be correctly expressed at all.
>
> (OC, § 37)

As this last passage indicates, Wittgenstein (early and late) maintains that philosophy itself is characterised in terms of its relationship to (and tendency toward) nonsense:

> The results of philosophy are the discovery of some piece of plain nonsense and the bumps that the understanding has got by running up against the limits of language.
>
> (PI, § 20)

Notoriously, the *Tractatus* concludes with the verdict that it is itself nonsense:

> My propositions serve as elucidations in the following way: anyone who understands me recognizes them as nonsensical, when he has used them—as steps—to climb beyond them.
>
> (TLP, 6.54)

Given the important role that the notion of nonsense plays throughout Wittgenstein's work, an adequate understanding of that work requires an understanding of what nonsense is, for Wittgenstein, and what makes for it.

To make progress on this front, Glock distinguishes two accounts of nonsense.[1] According to *the privation view*, represented by the opening passage from the *Notebooks*, nonsense results 'from our not having assigned a meaning to expressions in a certain context' (2004, 222). On this account, to borrow a well-worn example from Carnap (1959), the sentence 'Julius Caesar is a prime number' is nonsense because of a failure to give a meaning to its constituent expressions.[2] Although they might look like familiar English expressions, in this context one or more of them is, in fact, a meaningless sign.

According to *the combinatorial view*, in contrast, nonsense results from the way in which meaningful expressions are strung together or, as Glock puts it, 'from the illicit combination of meaningful words' (2004, 222). On this account, 'Julius Caesar is a prime number' is nonsense because, in a sense to be spelled out, the meanings of the words in the sentence do not allow for this combination.

'Why choose?' one might ask. Perhaps there are just different sorts of nonsense—the privation sort and the combinatorial sort. However,

there is an influential tradition in Wittgenstein scholarship according to which Wittgenstein—throughout his philosophical career—held only the privation view (Diamond 1991; Conant 1998, 2000; Crary 2000, 121; Witherspoon 2000; Mulhall 2007). Against this, Glock defends a pluralistic interpretation. According to it, Wittgenstein—early and late—holds that nonsense can occur as a result of privation *and* as a result of combination.[3]

In this chapter, I will argue that this is a mistake. In his later work, Wittgenstein does not endorse the combinatorial view or, for that matter, views that entail it. More cautiously, I will argue that the considerations that Glock adduces to support the attribution to Wittgenstein of the combinatorial view do not, in fact, support it. Does this mean that I attribute (only) the privation view to the later Wittgenstein? Yes, but, well, sort of.

In this debate, the term 'nonsense' is used in a somewhat restricted manner, specifically, to mean *unintelligible*. As Glock notes (2015, 120), in ordinary parlance, the term can be used instead to mean *pointless* or *futile*. In this sense, it might be nonsense for me to say, 'I am sorry' when it is mutual knowledge that I will perform the relevant action again without hesitation, though, of course, my sentence is not devoid of sense or unintelligible—it expresses a thought, albeit an insincere one. To anticipate, central to my reading of Wittgenstein is that, although he does not identify the two, he considers there to be a close connection between intelligibility and pointfulness.

To return to the issue at hand, I will argue that, for Wittgenstein, nonsense always results from a lack. However, what is lacking is not best described as the assignment of a meaning to an expression; instead, what is lacking is a purpose that the words might be understood as serving on the relevant occasion. At this stage, this is at best suggestive. I will explain the idea more fully in due course.

Before getting stuck in, there are two things to note. First, my concerns here are interpretive, not substantive. The question is what view Wittgenstein holds, not whether that view is true or well-supported.

Second, my claims are restricted to Wittgenstein's *later* philosophy. So, I do not deny that in his early or middle periods Wittgenstein advanced the combinatorial view (cp. PO, 58–59; VW, 235). My claim is that, if he did advance it then, he abandoned it in his later work (roughly, after 1935).

Given this focus, one might think that there is a quick argument for the pluralist reading: The view that there is something common to all cases of nonsense is against both the letter and the spirit of the later philosophy. For Wittgenstein, those cases will fall under the concept of nonsense in virtue of their position within 'a complicated network of similarities overlapping and criss-crossing', that is, in virtue of 'family resemblances' (PI, §§66–67). So, any reading according to which

Wittgenstein takes nonsense to consist always and only in one thing—privation—is a nonstarter.

The quick argument is too quick. Grant that there are many members to the nonsense family. One member of that family is *unintelligibility*. Concerning that family member, we can ask whether it results from combination or (only) from privation.

A proponent of the argument might say that this only postpones the issue. For the later Wittgenstein, unintelligibility too is a family resemblance concept.

No doubt. But there is still a debate to be had. Among the sentences that fall under the concept of being unintelligible, are there some, according to Wittgenstein, that do so because they involve meaningful words in illegitimate combinations? Glock says, 'Yes'. In what follows, I will explain why I am not yet convinced of that answer.

Meaning and use

There are two strategies for resolving this interpretive matter. The top-down approach is to explore Wittgenstein's later conception of meaning and ask whether it rules in or out combinatorial nonsense. The bottom-up approach is to look at specific passages and ask whether Wittgenstein there expresses a commitment to the possibility or otherwise of combinatorial nonsense. I will explore each strategy in turn, starting with the top-down approach.

The main argument that Glock identifies for attributing (only) the privation view to Wittgenstein appeals to Frege's (1953, § 62) context principle. Wittgenstein refers with approval to that principle, which he reformulates as follows:

> A word has meaning only as part of a sentence [with a sense].
> (PI, § 49)

The context principle, so understood, might seem to support the privation view and rule out the combinatorial view. Suppose a sentence lacks sense. It follows that its constituent words lack meaning. In turn, it follows that the sentence's lack of sense cannot be due to what those words mean, since they do not mean anything.[4]

However, as Glock argues, 'Wittgenstein's later conception of meaning militates [. . .] against the restrictive [context] principle' (2004, 228). According to Wittgenstein:

> For a *large* class of cases of the employment of the word 'meaning'—though not for all—this word can be explained in this way: the meaning of a word is its use in the language.
> (PI, § 43; see also § 432)

This might simply seem a more liberal version of the context principle, but it is clear that Wittgenstein takes the meaning of a word to

correspond, not to its use on a particular occasion, but to its usage, that is, to the way of using it.[5] Consider:

> Only in the practice of a language can a word have meaning.
> (RFM, VI § 41)

> A meaning of a word is a kind of employment of it. [. . .] That is why there exists a correspondence between the concepts 'rule' and 'meaning'.
> (OC, §§ 61–62)

The last remark points to a further aspect of Wittgenstein's conception of meaning. For a word to have meaning is for there to be rules governing its employment (cp. PI, § 355). He goes as far as to say:

> The rule-governed nature of our languages permeates our life.
> (ROC, § 303)

Given this conception of meaning, as determined by rules for the use of expressions, it is no surprise that Wittgenstein introduces the (infamous) notion of a language-game:

> Following according to the rule is FUNDAMENTAL to our language-game.
> (RFM, VI § 28)

This takes us back to Wittgenstein's apparent endorsement of the context principle. As he explains in PI, § 49, Wittgenstein takes the truth in that principle to be that a word is not a name, or more generally does not have a meaning, 'except in a game', that is, unless there is a rule-governed use for it.

This reading of the later Wittgenstein is familiar enough. The point is that, if correct, the argument from the context principle in support of the view that he recognises only nonsense by privation is unsound. Wittgenstein did not really think that a word has meaning only when it appears in a sentence with a sense. A word that does not appear in such a sentence might, nonetheless, have meaning insofar as there is a rule-governed use for it (Glock 2004, 229).

Moreover, one might think that Wittgenstein's later conception of meaning provides a straightforward explanation for how nonsense occurs other than by way of privation. If expressions are meaningful in virtue of being subject to rules of use, then nonsense results 'from combining meaningful expressions in a way that is prohibited by the rules for the use of these expressions' (Glock 2004, 222).

One response to this is to deny that Wittgenstein accepts the relevant conception of meaning. According to Cavell, for example, Wittgenstein's view is precisely that 'everyday language does not, in fact or in essence, depend upon such a structure [. . .] of rules' (1976, 48).[6] However, I

will grant that Wittgenstein thinks of meaning as a matter of rules of use but argue that he is anyway not committed to the combinatorial view.

This is instructive. Some interpreters who attribute to Wittgenstein (only) the privation view seem to think that, were Wittgenstein to hold that expressions have context-independent meanings in virtue of the rules governing their employment, then he would allow for combinatorial nonsense. For example, Witherspoon writes, 'If words belong to a system of [rule-governed] symbols, then they are meaningful symbols. This leaves open the possibility that meaningful symbols might be combined into a sentence-like formation that is not itself meaningful' (2000, 331). That possibility might be consistent with the view of meaning as consisting in rule-governed use, but it is not entailed by it, or so I will argue.

Sense and pointfulness

First, I will show that the reading of Wittgenstein sketched in the previous section, though correct, is incomplete. There is a further dimension to his later conception of meaning, one that is not overlooked in the literature, but that does not always receive the emphasis it deserves. According to Wittgenstein:

> The word 'language-*game*' is used here to emphasize the fact that the *speaking* of a language is part of an activity, or form of life.
> (PI, § 23; see also PI, § 19; ROC § 302)

Following Cavell (1989, 41; see also Moyal-Sharrock 2015), we might distinguish two senses of 'form of life'. There is a biological sense—concerning a life-form—and an ethnological sense—concerning a way of life. I focus here on the latter, which Wittgenstein has in mind in remarks such as the following:

> What belongs to a language-game is a whole culture.
> (LC, 8)

> Language, I should like to say, relates to a *way* of living.
> (RFM, VI § 34)

Insofar as a language-game is bound up with a form of life—in the ethnological sense—it is bound up with certain values, ends, or interests. To illustrate, consider:

> A tribe has two concepts, akin to our 'pain'. One is applied where there is visible damage and is linked with tending, pity, etc. The other is used for stomach-ache, for example, and is tied up with mockery of anyone who complains. 'But then do they really not notice the

> similarity?'—Do we have a single concept everywhere where there is similarity? The question is: Is the similarity *important* to them?
> (Z, § 380)

In this case, the fact that the two pain-like terms are to be used in different ways reflects what matters to the members of the tribe. Perhaps they have treatment only for tissue damage. Since expressions of internal pain serve no purpose and waste precious time and resources, they are to be discouraged. The point is not how to interpret the example—its details can be filled in in countless ways—but to illustrate that, for Wittgenstein, the rules of a language-game, hence, the meanings of expressions, are bound up with and in the service of certain ends or purposes:

> The game, one would like to say, has not only rules but also a *point*.
> (PI, § 564)

> Concepts lead us to make investigations. They are the expression of our interest and direct our interest.
> (PI, § 570)

> We could say that people's concepts show what matters to them and what doesn't.
> (ROC, § 293; see also RFM, I § 74; Z, §§ 387–388)

To sum up, Wittgenstein's later conception of meaning is one according to which a word has meaning in virtue of there being a usage for it, more specifically, a usage subject to rules that determine whether that word is employed correctly or incorrectly on a given occasion, more specifically still, rules bound up with the interests and needs of language-users. This conception points to a distinctive way of cashing out the privation view. For Wittgenstein, nonsense results when a person uses words in the absence of the sort of situation that provides or speaks to the relevant interests or needs (cp. RFM, I § 132).

By way of analogy, suppose that two people are moving what look like chess pieces around a chequered board. They move the horse-shaped pieces two squares in one direction, one in another. They move the castle-shaped pieces horizontally and vertically, never diagonally. When one piece enters the square another occupies, the latter piece is removed from the board. And so on. However, the goal of the activity is for each player to remove only the other's horse-shaped pieces. Or perhaps it is simply to make pretty patterns. Whatever the players are doing, it is not playing chess. Given the goals that inform the activity, the pieces as employed on these occasions are not subject to the rules of chess.

In a similar fashion, for Wittgenstein, nonsense results when a person utters words that might look like familiar words of (say) English, but

that utterance is not recognisable as in the service of the purposes with which the rules governing the relevant English words are bound up. So, whatever the person is doing in that situation, they are not using those English words that are subject to those rules.

To make this more concrete, consider:

> The words 'I am here' have a meaning only in certain contexts, and not when I say them to someone who is sitting in front of me and sees me clearly.
>
> (OC, § 348)

This passage is especially relevant for the present dispute, as the putative nonsense is clearly not of the combinatorial sort—Wittgenstein explicitly allows that that combination of words might express a sense in a suitable context. The issue, then, is that the context is not suitable. Why? Because there is no point or purpose with which the use of the relevant expressions engages. For illustrative purposes, suppose that the role of the indexicals 'I' and 'here' is to allow the audience to locate or identify a person and place, respectively. In the context that Wittgenstein describes, there is no such need to be met. So, whatever is going on there, it is not a context in which the speaker is participating in the practice of using expressions that answer to or serve that need.

By way of contrast, consider a context in which two people are having a conversation via Zoom. One holds a map up to the camera and says, 'I am here', perhaps adding, 'And our friend is there' (cp. Whiting 2017, 426–428).

It is helpful to compare this reading of Wittgenstein with one that Conant—an influential proponent of the privation (only) interpretation—rejects. According to it:

> When Wittgenstein says it [a sentence such as 'I am here'] is being used in an *unsuitable* situation, what he means is that we do not understand the *point* of the speaker's saying this perfectly determinate thing when he does.
>
> (1998, 230)

The view I attribute to Wittgenstein is not that 'I am here' makes sense in the context he describes, but its utterance serves no discernible purpose; rather, the view is that, when the utterance of 'I am here' serves no discernible purpose, the sentence in that context lacks sense.

Although he does not develop the point as I have done here, Glock agrees that in Wittgenstein's view:

> Whether an utterance makes sense, and what sense it makes, is not simply determined by the linguistic form of the sentence uttered, its

> constituents and mode of combination [. . .] Instead, it depends on the circumstances in which the utterance is made.
>
> (2004, 232)[7]

But Glock thinks this point speaks against, not for, the idea that nonsense results only via privation. If an expression lacks meaning '*because of its context* then its lack of meaning is not just a matter of privation, it is a matter of being in inappropriate company, just as the combinatorial view has it' (2004, 229–230; also 2015, 125). So, in attributing the above conception of what makes for nonsense to Wittgenstein, one might think that I concede the main point.

However, a better way to put the idea is that the expression is meaningless because *it lacks appropriate company*. The dearth of a suitable end or purpose amounts to the *absence* of an *enabling* condition on sense, rather than the *presence* of a *disabling* condition (cp. Bader 2016).

To put this another way, my claim is not that, for Wittgenstein, nonsense occurs due to some sort of 'clash' or 'incompatibility' between the sense of a sentence—or the meanings of its constituent expressions and the rules for their employment—and the context of utterance (cp. Conant 1998, 223), which might be understood as a matter of the circumstances frustrating or conflicting with the point or purpose of using those rule-governed expressions in that combination. Rather, my claim is that, for Wittgenstein, nonsense occurs when an utterance simply does not count as the use of expressions subject to the relevant rules because there is no connection between the circumstances of utterance and the point or purpose associated with the rule-governed use of those expressions.

The chess analogy is helpful here. The suggestion is not that, if people are moving pieces so as to make pretty patterns (etc.), they are playing chess in a way that violates its rules or frustrates its goal; it is rather that, given their goal, those people are not playing chess at all, hence, the context is not one to which the rules of chess apply.

One might object to an interpretation of Wittgenstein that relates rules governing the use of words—grammatical rules—to certain ends on the grounds that it clashes with his claim that such rules are 'arbitrary':

> Grammar does not tell us how language must be constructed in order to fulfil its purpose.
>
> (PI § 496; see also Z, § 322)

There are many dimensions to Wittgenstein's insistence on the arbitrariness of grammar (see Forster 2004). The salient one here is that, unlike the rules of cooking, the rules that constitute language-games are not justifiable or evaluable by reference to some independently specifiable purpose or goal. If I stick to certain instructions when cooking, I might cook badly. If I use others, I might cook well. So, in a straightforward

sense, the first set of instructions are worse than the second. Alternatively: So far as cooking goes, following the first set of instructions is wrong, whereas following the second set is right. In contrast, according to Wittgenstein, 'if you follow grammatical rules other than such-and-such ones, that does not mean you say something wrong, no, you are speaking of something else' (Z, § 320).

The view I attribute to Wittgenstein is not that the rules of grammar are to be justified or evaluated by reference to the relevant ends. It is, rather, that the rules are bound up with certain ends such that, if on a given occasion the use of a word is unconnected to those ends, then it does not qualify as a use of the word to which those rules apply, though it might qualify as the use of another word to which some other rules apply. So, I do not deny that, for Wittgenstein, the meaning-constituting rules for the use of words are in the relevant sense arbitrary.

Explanations of meaning

On the reading of Wittgenstein I have developed, nonsense is not the result of the meanings of the relevant expressions or, more carefully, the rules prohibiting their use in certain combinations. In support of this reading, I have insisted that we look at Wittgenstein's wider conception of meaning. But this cuts both ways. Glock argues that some of Wittgenstein's other commitments are incompatible with the view that sentences like 'Julius Caesar is a prime number' are nonsense due only to privation. In particular, Glock (2004, 230) invites us to consider Wittgenstein's suggestion that there is a correlation between the meaning of a term and an explanation of its meaning:

> If you want to understand the use of the word 'meaning', look for what one calls 'an explanation of meaning'.
>
> (PI, § 560)

With respect to Carnap's sentence, a person who utters it might explain, '"Julius Caesar" refers to the famous Roman general, and "is a prime number" means *is divisible only by 1 and itself*'. If the meaning of an expression is what is revealed in an explanation of its meaning, this person reveals meanings for the relevant expressions, despite the sentence involving them lacking sense.

This point proves too much. Consider its application to the Carnap sentence as a whole. Its utterer says, '"Julius Caesar is a prime number" means *the famous Roman general is divisible only by 1 and himself*'. Clearly, Wittgenstein would not take this to show that the sentence has a sense.

The thought behind PI, § 560 is that what a word means cannot transcend the explanations that speakers give of its meaning. As Glock is

aware, it does not follow from this that any explanation a person gives of a word captures what, if anything, it means. Suppose, for example, that a person says, by way of explanation, that the word 'cold' (in English) means *warm*. Given its established usage, this is false.

Another way to put this is to say that, for Wittgenstein, the meaning of a word corresponds to a *successful* explanation of it, i.e., one that succeeds in explaining what the word means. The dispute is then precisely whether the speaker's explanations of the words that appear in the Carnap sentence are successful.

I have argued that it is consistent with Wittgenstein's suggestion that meaning and explanation are coordinate notions that nonsense results only from privation (of purpose). However, reflection on such explanations leads to another argument against this interpretation.

Explanations of meaning serve, for Wittgenstein, as expressions of grammatical rules, such as:

1 'Julius Caesar' applies to the famous Roman general.
2 'is a prime number' applies (only) to a number that can be divided only by 1 and by itself.

One might think that the very idea that rules of this sort determine the meanings of the expressions they concern brings with it the idea of combinatorial nonsense. Indeed, Glock suggests that (1) and (2) entail that Carnap's sentence is nonsense:

> The explanations imply that the referent of 'Julius Caesar' is not within the range of meaningful application of 'is a prime number'.
> (2015, 124)

In that case, the sentence lacks a sense because of the rules governing its constituent expressions, hence, because of what they mean. In that case, in turn, there is combinatorial nonsense.

However, the rules do not imply what Glock says they do; rather, they imply that the referent of Julius Caesar is not within the range of application of 'is a prime number', that is, within its extension. So, what those rules imply is that Carnap's sentence is false, rather than nonsense, in any context of utterance in which its constituent expressions qualify as subject to (1) and (2).[8]

My reply to Glock assumes that 'applies to' as it occurs in (1) and (2) is equivalent to 'is true of'. In other work, Glock argues that the rules determinative of the meanings of expressions are not rules of truth of this sort. As he puts it, 'one can apply a word in a way which is *semantically correct*, without applying it correctly in the sense of applying it to say something *true*' (2005, 299, see also 1996a, 150–151, 2019). Elsewhere, I have argued against views of this sort on substantive grounds (Whiting

2016; Whiting Forthcoming). But, as stressed at the outset, the present focus is on exegetical matters.[9] It is time, then, to look at the passages that Glock appeals to in support of his reading and assess whether they do, in fact, support it.

A drop of grammar

I turn now to the bottom-up approach to defending the pluralist interpretation.

Glock quotes the following when defending the claim that, for Wittgenstein, the meaning-determining rules 'draw the bounds between correct and incorrect uses of words', such that incorrect usage results in nonsense (2004, 233):

> 'I know what I want, wish, believe, feel, [. . .]' (and so on through all the psychological verbs) is either philosophers' nonsense or, at any rate, *not* a judgment *a priori*.
>
> (PPF, § 309)

Also:

> I can know what someone else is thinking, not what I am thinking.
>
> It is correct to say 'I know what you are thinking', and wrong to say 'I know what I am thinking'.
>
> (A whole cloud of philosophy condenses into a drop of grammar.)
>
> (PPF, § 315)

Here is a way of capturing what I take to be Glock's reading of these remarks. Wittgenstein starts with some examples of philosophical nonsense ('I know' [. . .]). The remark that follows concerning what it is correct and incorrect to say is then the voice, not of the philosopher, but of Wittgenstein offering a diagnosis of the nonsense: The philosopher's words are nonsense, because they fail to accord with the rules governing them. This shows that, for Wittgenstein, nonsense results in some cases by combining expressions in ways that are prohibited by the rules for their employment.

There is, however, an alternative (I dare say, better) way to read the passages. Again, Wittgenstein starts with examples of the philosopher's nonsense. Why are they nonsense? Because their words do not engage with appropriate needs or interests. For illustrative purposes, suppose that the point of employing the English word 'know' is to resolve some doubt (OC, § 121) or to discriminate reliable from unreliable informants (OC, § 575).[10] The philosopher whom Wittgenstein invites us to imagine does not purport to speak to those needs when they use that word. So, their use of it is not a use of the English word 'know'.

The remark that follows concerning (in)correct use is not, I submit, Wittgenstein's explanation of why those words make no sense in that combination. Rather, it represents an attempt to find a sense for the philosopher's words, which requires considering 'the occasion and purpose of these phrases' (PPF, § 312). And Wittgenstein's suggestion is that we can imagine those words serving a purpose in ordinary—non-philosophical—contexts of linguistic instruction. So, the remark about what it is or is not correct to say is not in Wittgenstein's voice, but that of a person teaching someone that the word 'know' applies in the third-person case but not in the first-person case. In this way, I suggest, Wittgenstein is taking the words back from their 'philosophical' or 'metaphysical' use to their 'everyday' use (OC, § 347; PI, §§ 116, 372).

So, the passages from PPF do not support an interpretation of the later Wittgenstein according to which he thinks there is such a thing as combinatorial nonsense. Moreover, to return to an issue raised at the end of the last section, those passages do not demonstrate that Wittgenstein is using 'correct' to mean *semantically correct*, such that what it is correct to say comes apart from what it is true to say. It is consistent with PPF, § 315 that, in those contexts in which the rules are in force, hence, when the use of the expressions hooks up with relevant ends or concerns, 'knows' truly applies in the third-person case but not the first-person case.

The bounds of sense

Glock cites another series of passages from the later work as evidence that, for Wittgenstein, 'whether the occurrence of a word on a particular occasion results in nonsense depends at least partly on what other words it is combined with' (2004, 233; see also Schönbaumsfeld 2010, 655–656):

> When I say that the orders 'Bring me sugar!' and 'Bring me milk!' have a sense, but not the combination 'Milk me sugar', this does not mean that the utterance of this combination of words has no effect. And if its effect is that the other person stares at me and gapes, I don't on that account call it an order to stare at me and gape, even if that was precisely the effect that I wanted to produce.
>
> To say 'This combination of words has no sense' excludes it from the sphere of language, and thereby bounds the domain of language. But when one draws a boundary, it may be for various kinds of reason. If I surround an area with a fence or a line or otherwise, the purpose may be to prevent someone from getting in or out; but it may also be part of a game and the players are supposed, say, to jump over the boundary; or it may show where the property of one

> person ends and that of another begins; and so on. So if I draw a boundary-line, that is not yet to say what I am drawing it for.
>
> When a sentence is called senseless, it is not as it were its sense that is senseless. Rather, a combination of words is being excluded from the language, withdrawn from circulation.
>
> (PI, §§ 498–500)[11]

Why might these remarks be thought to commit Wittgenstein to the possibility of combinatorial nonsense? Perhaps because Wittgenstein here says that what has or lacks sense is a 'combination' of words. However, to refer to combinations of words may be just another way of referring to sentences, that is, to the items that are candidates for being sense and nonsense. Moreover, although Wittgenstein refers here to combinations of words, that is, to sentences, there is no suggestion that those combinations lack sense *in virtue of* the meanings of the words they contain, the ways in which they are combined, or the rules governing their use.

Wittgenstein also speaks in these passages of the 'boundaries of language'. Does that support the pluralist interpretation? It does if Wittgenstein thinks that the boundaries are fixed by rules that determine which combinations of words make sense and which do not. But, first, Wittgenstein does not say anything in the passages that commit him to this. Second, the privation view, at least, as I have developed it, points to a different way in which a boundary to language might be determined—it lies at the point at which the use of the relevant expressions ceases to engage with suitable interests or purposes.

One might object that my version of the privation view struggles to make sense of Wittgenstein's idea of *exclusion from a language*. On the interpretation I defend, a combination of words such as 'I am here' that in one context lacks sense might in another context have sense, hence, belong to a language. In that case, the combination is not really excluded.

If there is a problem here, it is a problem, not only for my reading of Wittgenstein, but also for Glock's pluralist alternative. Glock agrees that, for Wittgenstein, nonsense can result from the absence of a suitable context, hence, that sense can be secured by the provision of a suitable context. In any event, there is no problem here. Exclusions need not be absolute. Indeed, Wittgenstein's remarks on the different purposes for which a boundary to language might be drawn suggest that they are not.

So, the interpretation I defend has no difficulty with the idea of exclusion from a language. Moreover, I suggest that it does a better job of capturing that idea than its competitor. On the combinatorial view, although 'Julius Caesar is a prime number' is nonsense, it is constituted by words of English. In that case, it is an English sentence, hence, a sentence in a language. On the privation view, in contrast, although those words might look like words of the English language, they are not.

It is instructive to note that the themes of exclusion and senseless sense that run throughout PI, §§498–500 crop up again in a remark from the same period concerning Moorean sentences of the form: *p*, but I do not believe that *p*.[12] Wittgenstein writes:

> Again, you must not forget that 'A contradiction doesn't make sense' does not mean that the sense of a contradiction is nonsense.—We exclude contradictions from our language; we have no clear-cut use for them, and we don't want to use them. And if 'It's raining but I don't believe it' is senseless, then again that is because an extension along certain lines leads to this technique. But under unusual circumstances that sentence could be given a clear sense.
>
> (RPP II, § 290)

Here, Wittgenstein allows that in a suitable context the Moorean sentence has a sense. What context? Wittgenstein is not explicit about this, but he does offer some clues. He suggests that the sentence would make sense if the speech in which it figures were 'automatic' (RPP II, § 292). To explain what he means by this, Wittgenstein invites us to imagine a case in which '*two* people are talking through one mouth' (RPP II, § 293). Presumably, one person utters the first conjunct—'It is raining'—whereas the other utters the second—'I don't believe it'.

I take it that what Wittgenstein is thinking of here is a situation in which a person's mind is fragmentated and in which they are alienated or disassociated from their own thoughts. Consider the hackneyed example of the patient on the therapist's couch: My parents love me, but I don't believe that. If the Moorean sentence makes sense in a therapeutic context, or some other circumstance of dissociation, then its senselessness outside of such a context cannot be due to the fact that the rules governing the use of its constituent expressions prohibit that combination. After all, those rules do allow for those expressions in that combination to express a sense.

To return in light of this to the issue at hand: In his remarks on the theme of senseless sense in RPP II, Wittgenstein does not have in mind combinatorial nonsense. This supports the suggestion that, in the companion remarks on the same theme in PI, Wittgenstein does not (or, at the very least, need not) have in mind combinatorial nonsense. Moreover, Wittgenstein's remarks on the Moorean sentence confirm that he views the exclusion of a combination of words from a language as neither permanent nor unconditional.

Conclusion

I have defended an interpretation of Wittgenstein's later philosophy according to which nonsense results only from privation. When a sentence

is senseless, on Wittgenstein's view, what is lacking is not so much the assignment of meanings to its constituent expressions—an idea that smacks of some kind of semantic voluntarism—but, rather, a point or purpose with which its words engage.

This reading, I have argued, is not in tension with attributing to Wittgenstein the view that words have context-invariant meanings in virtue of rules governing their use. The key idea is that those rules are themselves bound up with certain needs or ends such that, if a context of use is not related to those ends, it is not a context to which those rules apply.

In contrast, Glock defends an interpretation according to which, for Wittgenstein, nonsense (also) results from using words in ways that the rules for their use prohibit. I have argued that the passages Glock cites in support of this reading do not, in fact, support it, and also that a commitment to such combinatorial nonsense is not a consequence of Wittgenstein's later conception of meaning.

Some defenders of a privation reading suggest that, for Wittgenstein, there is no significant difference between philosophers' nonsense—like 'Julius Caesar is a prime number', 'I know what I am thinking', etc.—and mere gibberish (Conant 2001, 14; Diamond 2000, 151). Nothing I have said here commits Wittgenstein to such a view. What difference there is between philosophers' nonsense and mere gibberish, for Wittgenstein, is a nice question, but it is a question for another occasion.[13]

Works by Wittgenstein

I use the standard abbreviations to refer to Ludwig's Wittgenstein's works:

LC *Lectures and Conversations on Aesthetics, Psychology, and Religious Belief*. Edited by C. Barrett. 1966. Oxford: Blackwell.

NB *Notebooks 1914–16*. 2nd edition. Edited by G.E.M. Anscombe and P. Geach. Translated by G.E.M. Anscombe. Oxford: Blackwell, 1979.

OC *On Certainty*. Edited by G.E.M. Anscombe and G.H. von Wright. Translated by D. Paul and G.E.M. Anscombe. Oxford: Blackwell, 1969.

PO *Philosophical Occasions: 1912–1951*. Edited by J. Klagge and A. Nordmann. Indianapolis, IN: Hackett, 1993.

PI *Philosophical Investigations*. Edited by G.E. M. Anscombe and R. Rhees, revised by P.M.S. Hacker and J. Schulte. Translated by G.E.M. Anscombe, P.M.S. Hacker and J. Schulte. Oxford: Blackwell, 1953/2009.

PPF *Philosophy of Psychology—A Fragment*. In PI, 183–243.

RFM *Remarks on the Foundations of Mathematics*. 3rd edition. Edited by G.H. von Wright, R. Rhees and G.E.M. Anscombe. Translated by G.E.M. Anscombe. Oxford: Blackwell, 1978.

ROC *Remarks on Colour*. Edited by G.E.M. Anscombe. Translated by L.L. McAlister and M. Schättle. Oxford: Blackwell, 1980.

RPP II *Remarks on the Philosophy of Psychology*. Vol. II. Edited by G.H. von Wright and H. Nyman. Translated by C.G. Luckhardt and M.A.E. Aue. Oxford: Blackwell, 1980.

TLP *Tractatus-Logico Philosophicus*. Translated by D.F. Pears and B.F. McGuinness. London: Routledge, 1961.

VW *The Voices of Wittgenstein: The Vienna Circle, Ludwig Wittgenstein, and Friedrich Waismann*. Edited by G.P. Baker. Translated by G.P. Baker, M. Mackert, J. Connolly and V. Politis. London: Routledge, 2003.

Z *Zettel*. Edited by G.E.M. Anscombe and G.H. von Wright. Translated by G.E.M. Anscombe. Oxford: Blackwell, 1967.

Notes

* For feedback on earlier versions of this material, I am grateful to Denis McManus, the editors of this volume, and audience members at the University of Zürich, especially Hanoch Ben-Yami, Severin Schroeder, and Hanjo Glock.

1 For the same distinction in different terms, see Diamond (1991) and Conant (2001).

2 For the sake of argument, I will take it for granted that such sentences are (in the sorts of context we are invited to imagine) nonsense, as opposed to (merely) false. For discussion, see Magidor (2016).

3 For influential elaborations and defences of this reading, see Baker and Hacker (2009), Hacker (1986). See also Glock (1996a, 260–264. 1996c, 184–185).

4 For this line of thought in relation to Wittgenstein's *Tractatus*, see Conant (2000, 191–194) and Diamond (1991, 98–100). For the same line of thought in relation to the later Wittgenstein, see Diamond (1991, 107) and Witherspoon (2000, 323–324).

5 For this point, see Baker and Hacker (2005, 153), Glock (1996b, 207), Schroeder (2006, 172), and Whiting (2008).

6 For otherwise diverse readings of Wittgenstein that deny he takes language to involve rules, see Glüer and Wikforss (2010), Hanfling (1980), Hutchinson (2007), Luntley (2003), and Witherspoon (2000).

7 A further claim is that, for Wittgenstein, the same sentence in different contexts might bear a different sense (see Conant 1998; Dobler 2013; Travis 1989, 2006). For challenges to this interpretation, see Bridges (2000) and Whiting (2017).

8 To address this, Glock might strengthen (2) as follows:

> (3) 'is a prime number' meaningfully applies (only) to any number that can be divided only by 1 and by itself.

However, (3) is too strong. With (1), it entails that the following sentence is senseless: The number of countries in the United Kingdom is a prime number.

9 Moore suggests that Wittgenstein has such a distinction in mind but on the basis of lectures from the 'middle period' (PO, 79–80), so I set this aside. In his later work, Wittgenstein suggests that grammatical rules include those governing the substitution of expressions or inference (e.g., PI, §558). One might call these rules for *truth-preservation*.

10 For more on this theme, see Craig (1990).

11 Diamond (1991, 106–109) and Conant (2001, 14) suggest that PI, §500 rules out the possibility of combinatorial nonsense. For a response to this, see Schönbaumsfeld (2010, 655–656).

12 For different perspectives on Wittgenstein on Moore's paradox, see Heal (1994), Williams (1998), and Moran (2002).

13 For a discussion, see McManus (2014).

References

Bader, R. (2016). Conditions, Modifiers, and Holism. In Lord, E. and Maguire, B. (eds.), *Weighing Reasons*. Oxford: Oxford University Press, pp. 27–55.

Baker, G.P. and Hacker, P.M.S. (2005). *Wittgenstein: Understanding and Meaning*, revised edition. Oxford: Blackwell.

Baker, G.P. and Hacker, P.M.S. (2009). *Wittgenstein: Rules, Grammar, and Necessity*. Rev. Oxford: Blackwell.

Bridges, J. (2000). Wittgenstein vs. Contextualism. In Ahmed, A. (ed.), *Wittgenstein's Philosophical Investigations*. Cambridge: Cambridge University Press, pp. 109–128.

Carnap, R. (1959). The Elimination of Metaphysics through the Logical Analysis of Language. In Ayer, A.J. (ed.), *Logical Positivism*. New York: Free Press, pp. 60–81.

Cavell, S. (1976). *Must We Mean What We Say*. Cambridge: Cambridge University Press.

Cavell, S. (1989). *His New Yet Unapproachable America*. Chicago, IL: Chicago University Press.

Conant, J. (1998). Wittgenstein on Meaning and Use. *Philosophical Investigations* 21: 222–250.

Conant, J. (2000). Elucidation and Nonsense in Frege and Early Wittgenstein. In Crary, A. and Read, R. (eds.), *The New Wittgenstein*. London: Routledge, pp. 174–217.

Conant, J. (2001). Two Conceptions of Die Überwindung der Metaphysik. In McCarthy, T. and Stidd, S. (eds.), *Wittgenstein in America*. Oxford: Oxford University Press, pp. 13–61.

Craig, E. (1990). *Knowledge and the State of Nature*. Oxford: Oxford University Press.

Crary, A. (2000). Wittgenstein's Philosophy in Relation to Political Thought. In Crary, A. and Read, R. (eds.), *The New Wittgenstein*. London: Routledge, pp. 118–146.

Diamond, C. (1991). *The Realistic Spirit*. Cambridge: MIT Press.

Diamond, C. (2000). Ethics, Imagination and the Method of Wittgenstein's *Tractatus*. In Crary, A. and Read, R. (eds.), *The New Wittgenstein*. London: Routledge, pp. 149–173.

Dobler, T. (2013). What Is Wrong with Hacker's Wittgenstein? *Philosophical Investigations* 36: 231–250.
Forster, M.N. (2004). *Wittgenstein on the Arbitrariness of Grammar*. Princeton, NJ: Princeton University Press.
Frege, G. (1953). *The Foundations of Arithmetic*. 2nd edition. Translated by J.L. Austin. Evanston, IL: Northwestern University Press.
Glock, H.-J. (1996a). *A Wittgenstein Dictionary*. Oxford: Blackwell.
Glock, H.-J. (1996b). Abusing Use. *Dialectica* 50: 205–224.
Glock, H.-J. (1996c). Necessity and Normativity. In Sluga, H. and Stern, D. (eds.), *The Cambridge Companion to Wittgenstein*. Cambridge: Cambridge University Press, pp. 180–208.
Glock, H.-J. (2004). All Kinds of Nonsense. In Ammereller, E. and Fischer, E. (eds.), *Wittgenstein at Work*. London: Routledge, pp. 221–245.
Glock, H.-J. (2005). The Normativity of Meaning Made Simple. In Beckermann, A. and Nimtz, C. (eds.), *Philosophy—Science—Scientific Philosophy*. Paderborn: Mentis, pp. 199–241.
Glock, H.-J. (2015). Nonsense Made Intelligible. *Erkenntnis* 80: 111–136.
Glock, H.-J. (2019). The Normativity of Meaning Revisited. In Bayertz, K. and Roughley, N. (eds.), *The Normative Animal?* Oxford: Oxford University Press, pp. 295–318.
Glüer, K. and Wikforss, Å. (2010). Es braucht die Regel nicht. In Whiting, D. (ed.), *The Later Wittgenstein on Language*. Basingstoke: Palgrave, pp. 148–166.
Hacker, P.M.S. (1986). *Insight and Illusion*. 2nd edition. Oxford: Oxford University Press.
Hanfling, O. (1980). Does Language Need Rules? *Philosophical Quarterly* 30: 193–205.
Heal, J. (1994). Moore's Paradox. *Mind* 103: 5–24.
Hutchinson, P. (2007). What's the Point of Elucidation? *Metaphilosophy* 38: 691–713.
Luntley, M. (2003). *Wittgenstein: Meaning and Judgement*. Oxford: Blackwell.
Magidor, O. (2016). *Category Mistakes*. Oxford: Oxford University Press.
McManus, D. (2014). Austerity, Psychology, and the Intelligibility of Nonsense. *Philosophical Topics* 42: 161–199.
Moran, R. (2002). *Authority and Estrangement*. Princeton, NJ: Princeton University Press.
Moyal-Sharrock, D. (2015). Wittgenstein on Forms of Life, Patterns of Life, and Ways of Living. *Nordic Wittgenstein Review* 4: 21–42.
Mulhall, S. (2007). *Wittgenstein's Private Language*. Oxford: Oxford University Press.
Schönbaumsfeld, G. (2010). A 'Resolute' Later Wittgenstein? *Metaphilosophy* 41: 649–668.
Schroeder, S. (2006). *Wittgenstein*. Cambridge: Polity.
Travis, C. (1989). *The Uses of Sense*. Oxford: Oxford University Press.
Travis, C. (2006). *Thought's Footing*. Oxford: Oxford University Press.
Whiting, D. (2008). The Use of 'Use'. *Grazer Philosophische Studien* 76: 135–147.
Whiting, D. (2016). What Is the Normativity of Meaning? *Inquiry* 59: 219–238.
Whiting, D. (2017). Languages, Language-Games, and Forms of Life. In Glock, H.-J. and Hyman, J. (eds.), *A Companion to Wittgenstein*. Oxford: Wiley Blackwell, pp. 420–432.

Whiting, D. (Forthcoming). Semantic Normativity, Properly So Called. In Verheggen, C. (ed.), *Kripke's Wittgenstein on Rules and Private Language at 40*. Cambridge: Cambridge University Press. p. 58.

Williams, J.N. (1998). Wittgensteinian Accounts of Moorean Absurdity. *Philosophical Studies* 93: 283–306.

Witherspoon, E. (2000). Conceptions of Nonsense in Carnap and Wittgenstein. In Crary, A. and Read, R. (eds.), *The New Wittgenstein*. London: Routledge, pp. 315–349.

4 On safari with Glock

Constantine Sandis

> BÊTES: Ah! si les bêtes pouvaient parler! Il y en a qui sont plus intelligentes que des hommes.
>
> —Flaubert, Le *dictionnaire des idées reçues*

Prologue

Disclaimer: I have never been on safari with Professor Glock. His work on Wittgenstein, however, is the main reason I made the journey from the parklands of Oxford to the habitat of the Reading floodplains. I spent many—on some accounts, *too* many—happy years there as a graduate student in the University's philosophy department. Ever since those salad days, I have learned more from Hanjo than I could possibly express here; my gratitude to him is immense. Whatever criticism follows is intended to honour him, both as the author of some of my favourite philosophical texts and as the spirited conversationalist who can pull an argument apart whilst simultaneously scanning the restaurant table for any untouched desserts.

In 1996, my undergraduate self stepped into the basement of Blackwell's flagship bookshop on Broad Street and came out with a copy of Glock's *A Wittgenstein Dictionary* (1996a), priced at £10.99. It was within its pages that I first encountered the ideas I discuss in this chapter. Towards the end of his entry for 'form of life (*Lebensform*)', Glock offers a brief exegesis of Wittgenstein's 'puzzling remark' that 'if a lion could talk, we could not understand him' (Wittgenstein PPF, § 327).[1] At the time, I assumed that Glock's treatment of it was an undisputable orthodoxy. This was a *dictionary* after all, and I intended to use it to get through my final examinations. Some years later, I read his masterful paper 'On Safari with Wittgenstein, Quine and Davidson' (1996b), which was published around the same time as the *Dictionary*. The paper is a model of what philosophical writing can and should be: learned yet original, resolute but judicious, significant while light-hearted; its insights are always perspicuous and its reprovals constructive. Upon reading it, I also realised that far from being standard, Glock's account of

DOI: 10.4324/9781003202929-6

Wittgenstein was novel and important, both as an act of Wittgensteinian exegesis and as a piece of contemporary philosophical criticism.

Some years down the line, I nonetheless find myself at some distance from Glock's interpretation of the lion remark. This is a little awkward, for we both approach Wittgenstein's aphorism from a point of substantial agreement across an unlikely number of shared interests: Wittgenstein's later work, the nature of (human and non-human) animal thoughts and concepts, and philosophical anthropology at large (see e.g. Glock 2000, 2006). My hope in this essay is to try to convince Hanjo that his own understanding of Wittgenstein's ethnological method is best served by a different reading of the lion remark. I expect that my chances of success are not high. Thankfully, this is not a finals exam paper and Hanjo will not be grading it (at least not in print). I look forward to his undoubtedly instructive and illuminating reply.

Glock's lion

All but the final section of Glock's 'On Safari' is largely focused on explicating W.V. Quine's 'radical translation' and Donald Davidson's 'radical interpretation', with Wittgenstein only appearing as a third wheel of comparison. It is only in the paper's final section that Glock puts forth an interpretation of Wittgenstein's technical concept of a form of life as 'a culture or social formation' into which language-games are embedded (Glock 1996b, 165). This final section is the focus of my chapter. But whereas Glock appeals to the concept with the specific aim of building a powerful alternative to the philosophies of Quine and Davidson, I am only concerned with his interpretation of the Wittgenstein. Quine and Davidson have been booted out of the safari and replaced by Glock and Sandis.

I shall follow Glock (1996b, 170 n.1) in using 'radical translation' to refer simply to any attempt at understanding others that operates 'from scratch'. Both the scratchiness and understanding in question are, however, a matter of degree. Indeed, both Glock and I read Wittgenstein as demonstrating that we always have something in common with any being we might begin to understand. Strictly speaking, then, there is no such thing as *fully* radical translation; it is just that some translations are more radical than others. My quibble with Glock concerns the radicality of translating the talking lion. This partly hangs on the question of what it might even *be* for a lion to talk, a question that is inextricably tied to what Wittgenstein meant his aphorism to convey. As one of his students puts it in Derek Jarman's film *Wittgenstein* (1993), 'if we could understand him [Wittgenstein], I shouldn't think we'd have too much trouble with a lion'.

According to Glock, Wittgenstein's remark is best explained in light of the fact that 'we could never start the hermeneutical process unless "we"

shared with the interpretees certain *forms or facts* of life' (Glock 1996b, 166, Glock's emphasis). This is because 'understanding an alien language presupposes convergence not of beliefs, but of patterns of behaviour, which presuppose common perceptual capacities, needs and emotions' (Glock 1996a, 128). On both these points, we are in full agreement. Indeed, my niggle presupposes a shared understanding of the nature of forms of life and their relation to radical translation.

Glock offers two possible readings of Wittgenstein's remark about the talking lion, each resting on a different understanding of what he means by 'sprechen':

> On one reading, this means that we could not understand a lion who utters English sentences like 'I'm not interested in you, I've just eaten an antelope', which is obviously false (although one might, following Austin, question whether such a talkative creature could count as a lion). On a charitable reading, it means that if lions had a *feline* language of complex growls, roars, etc., we could never come to *learn* it. Why? Because their form of life, and their behavioural repertoire are so alien to us. We could not make head or tail of their facial expressions, gestures and demeanour. Moreover, our ability to interact even with a tame lion is strictly limited. For related reasons we 'could not find our feet' with human beings who give no expression of feeling of any kind, and would be completely at a loss with spherical Martians.
>
> (Glock 1996a, 166; see also 1996b, 128)

On the first interpretation that Glock outlines, the lion remark effectively means that we could not understand a lion who *speaks* by uttering sentences in human languages such as English, German, or Swahili. On the second interpretation, the term 'talk' would seem more apposite. People far more readily allow that animals *talk* to one another than that they *speak*, a fact which favours the 'speak' translation of 'sprechen', given Wittgenstein's implication that lions are not able to perform the activity in question.[2]

Glock claims that the second reading is the more charitable because the first 'is obviously false'. By contrast, I wish to argue that Glock presents us with a false dilemma. The question, I believe, is not whether the lion would talk in growls or words (including ones that sound just like German or English) but whether its lifeworld overlaps with that of humans in a way that would allow for a degree of shared concepts sufficient to make it possible for them to come to learn its language. How can we be sure that a speaking lion would say things like 'I've just eaten an antelope', even if it did utter sentences? Unlike the talking animals in Hollywood films and cartoons, wild animals do not share most of 'our' concepts.[3]

What about the first reading? First, a creature that speaks in words need not be speaking a *human* language. Glock's contrast between human (word-based) and non-human (growl) communication ignores the possibility of a word-based language that is not an actual human one that the relevant native speakers would understand. This could be the case even if, by some cosmic accident, its words sounded *just like* those of an actual human language. We need to further distinguish between what the expression 'I'm not interested in you, I've just eaten an antelope' would in this context mean in English and what it might mean in Lionese.

Given that words and sentences in our own languages can completely change their meaning across time and place (see Sandis 2016), we should not expect the imaginary lion's spoken words to mean what a native speaker might instinctively take them to mean. If they did so by stipulation, we would not be dealing with a talking lion at all but with a human 'in the shape of a lion' (Baker and Hacker 1985, 186; cf. Dennett 1979, 1991, 1995, 1996), for no other creature would have a grasp of the required concepts (cf. Schroeder 2019, § 4). It is through its behaviour and its behaviour alone that we could ever come to figure out what it might be saying, be it in words or growls.[4]

Talk is cheap

Wittgenstein is not concerned with linguistic comprehension but with interpersonal understanding. His point is not that we could not understand what the lion *says* but that we could not understand *the lion*. After all, we often understand the words someone utters while completely failing to understand the person.[5] As Wittgenstein puts in a remark preceding his quip about the speaking lion:

> We also say of a person that he is transparent to us. It is, however, important as regards our considerations that one human being can be a complete enigma to another. One learns this when one comes into a strange country with entirely strange traditions; and, what is more, even though one has mastered the country's language. One does not understand the people. (*And not because of not knowing what they are saying to themselves*.) We can't find our feet with them.
>
> (Wittgenstein 2009, § 325, my emphasis)[6]

Likewise, even if we did understand *what* the lion says (e.g., if it were speaking in English, German, or Swahili), that does not mean that we would be able to understand *why* it would say such a thing, whether its speaker meaning coincided with the conventional meaning of the sentences it utters, and so forth. This is all assuming it is even possible to

fully understand *what* someone is doing or saying without understanding *why* (cf. Anscombe 1957).

To complicate things more, we would (as with any human) need to be able to tell whether the lion was joking or being ironic, sarcastic, metaphorical, and so forth. We can do this only if we already share the relevant behavioural practices with it (e.g., in terms of gesture and tone) (see Glock 1996b, 166). But 'we' do not share a form of life with the lion; we simply do not have enough in common. And, indeed, as Wittgenstein's awkward remarks about 'the Chinese', 'English women', and 'Europeans' indicate,[7] 'we' does not stand for all humans but only for a contextually defined subset, one that would not obviously include lion tamers, let alone a Mowgli or a Tarzan (Sandis 2019a; cf. Schroeder 2019). While Glock is right to object that Wittgenstein 'is sometimes too impressed by the cultural diversity of gestures' (Glock 1996b, 167), he fails to note that there is no fixed 'we' which could or could not understand the speaking lion. Rather, 'we' functions as a kind of shifter in Roman Jakobson's sense, according to which the pronoun's referent 'shifts' depending on who is uttering it and in what context (Jakobson 1957; cf. Kursell 2010).

The same point may also be approached from the opposite direction. The talking animals in, say, *The Jungle Book* or *The Lion King* are in many ways far closer to human animals than they are to the non-human ones they are presented as. They say things like 'good luck', 'he has no right', or 'it is unsportsmanlike'; they talk of 'taking orders', 'speaking the truth', 'buying for a price', 'miracles', and 'treasure'. Yet none of these expressions involve concepts that an animal like a lion or tiger could have.[8] Such fictions basically portray humans in animal bodies. What prevents real-life lions from saying such things is not the inability to speak but a conceptual lack whose causes are largely biological. Non-human animals can be trained to share human practices only up to a point.

I am, thus, in sympathy with what Eike von Savigny (1991, 111) refers to as 'the received interpretation' (e.g. Pitcher 1964, 243), which is not so different from the one rejected by Glock in *A Wittgenstein Dictionary:*

> Let a lion be a competent speaker of some natural human language, competent in the sense preferred by linguists, namely, able to utter syntactically and semantically non-deviant sentences *ad libitum*. His utterances could not possibly be embedded in situations, reactions, and responses such that we could make sense of them. We should not be able to understand why he makes the utterances he makes or why he reacts to our utterances in this or that way. 'Good morning' would not come out when it would be appropriate for a greeting; 'thank you' might come as a reaction to a lash from a whip, and so on. These sentences, though perfectly good English, would

> not be used in the behavioural activities of greeting and thanking, and the lion in uttering them would not greet or thank someone. Hence these sentences would not have their standard English meanings, and the same would hold for any such utterances in so far as the lion does not share a form of life in which there are places for utterances with such meanings. In short: the lion does not share our form of life, and this is why we could not interpret anything he produces as an utterance that is translatable into a natural human language.
>
> (von Savigny 1991, 111–112)

At the very least, it seems to me correct to say that 'we' are not able to understand the communicative intention and speaker meaning of the lion precisely because our forms of life are sufficiently different to render the lion's reasons opaque. Von Savigny criticises such suggestions for taking the lion remark out of context (von Savigny 1991, 111).[9] But that is not the case if Wittgenstein intended the remark to contrast with the remarks that immediately precede it.

I am here in agreement with Simon Glendinning, although we disagree on the precise nature of the contrast between Wittgenstein's human-to-human scenario, in which we *do not* understand the people of a strange country and *cannot* find our feet with them (Wittgenstein 2019, § 325), and the human-to-lion scenario, in which we *could not* (or are not able to) understand the lion (Wittgenstein 2019, § 327, my emphasis).[10] Glendinning suggests that Wittgenstein is here stressing a difference of *order* in the opacity that a human being 'can encounter in its respective relations with other human beings and other animals' (Glendining 1998, 71). While we cannot find our feet with the strange humans, we presumably *could* do certain things that would render us capable. This suggests a failure to exercise an ability we already have. But even if we leave aside the contentious presumption that Wittgenstein's lion somehow stands for all non-human animals (see Sandis 2019a), the contrast does not seem as strong as Glendinning suggests. For it is not impossible for 'us' to *acquire* the ability to understand the speaking lion (see Sandis 2012, 154–155). Indeed, this predicament is similar to the one we face when we encounter the human tribe in Zettel, which 'we' are told is in some ways more alien to 'us' than a dog:

> Imagine that the people of a tribe were brought up from early youth to give no expression of feeling *of any kind* [. . .]. An education quite different from ours might also be the foundation for quite different concepts [. . .]. What interests us would not interest *them* [. . .]. 'These men would have nothing human about them'. Why? We could not possibly make ourselves understood to them. Not even as we can to a dog. We could not ['könnten nicht'] find our feet with

> them. And yet there surely could be such beings, who in other respects were human.
>
> (Wittgenstein Z, §§ 383–390)

Von Savigny contends that the only thing that matters for Wittgenstein's argument is that we *could not* understand the lion. Accordingly, 'further explanations for why we could not understand him are unnecessary in this context; the lion example might as well be read with a supplement: "Why—if a lion could talk, we could not understand him either"' (von Savigny 1991, 113). In opposition to this sort of approach, I have been trying to suggest that the lion remark is the conclusion of an argument that precedes it. If that is right, then one cannot fully appreciate the remark's function without a proper understanding of the argument that leads up to it. While I therefore agree with von Savigny that the immediate context is key, we disagree about what the context reveals.

Von Savigny, Glock, and I nevertheless all agree on a salient point, namely that neither the remark itself nor its wider context suggest that it is impossible to envisage mastering a language that is embedded in an alien form of life. On the contrary, Wittgenstein explicitly envisages such a scenario. According to von Savigny, it is sufficient that the behaviour of the objects of radical translation share certain regularities, regardless of whether the interpreting subject shares any of them. But this cannot be right either as an interpretation of Wittgenstein or as a hermeneutical truth. I, thus, concur with Glock's statement that

> like Quine and Davidson, Wittgenstein insists there are minimum requirements which a form of linguistic behaviour must meet in order to be intelligible to us. Our form of life need not be identical with that of the natives; after all, even if we leave aside Wittgenstein's fictional cases, we have managed to translate very remote languages such as Linear B, and to interpret very alien cultures, like that of the New Guinea head-hunters. But we could never start the hermeneutical process unless we shared with the interpretees certain forms or facts of life.
>
> (Glock 1996b, 166)[11]

As we shall see, however, while sharing these forms of life is crucial, it is not enough. What is shared is but the springboard for the acquisition of new abilities without which we would remain *incapable* of the relevant understanding.

Radical translation and forms of life

Pace Quine and Davidson, radical translation is not a question of theory construction but of 'being introduced into a normative practice' (Glock

1996b, 164). To enter another's 'hermeneutical circle', we need some kind of access to the 'cultural and social formation' that constitutes their 'form of life' (Glock 1996b, 165; see also von Savigny 1991, 109). The term 'form of life' occurs just six times in Wittgenstein's published output. As Glock notes, he adopted it to refer to 'the intertwining of culture, world-view, and language', including 'patterns of behaviour' (Glock 1996a, 124–125).[12] Failure to understand another frequently results from the lack of a shared form of life. At the very least, one must have enough in common with others so as to be able to find one's feet with them (a task that is impossible in the case of Glock's beloved spherical Martians). In *Nachlass* MS 124 (208ff.), in a remark preceding that which occurs in *Philosophical Investigations* § 206, Wittgenstein (n.d.) writes:

> I come to an alien people and someone apparently gives an order in a language which I do not know; his gestures, voice and the situation suggest to me that it is an order. I hear these sounds or words from different people in different circumstances expressed in the same tone of voice, But I see no regularity in the reactions of the other to whom the words are directed. Would I call these orders?
>
> In the reactions to an order, there must be uniformity.[13]

The activities that we perceive do not seem to be governed by any rules. They also do not form evidence for the existence of a language which is private to the alien people, for we would only be inclined to attribute them with language to the extent that their behaviour contains observable regularities. It is possible, of course, for these to occur in a fashion which renders them biologically (as opposed to metaphysically) inaccessible to us, though these can sometimes be overcome through the use of technology (e.g., infared vision goggles). Either way, the *criteria* for the possibility of understanding are ultimately behavioural:

> Suppose you came as an explorer to an unknown country with a language quite unknown to you. In what circumstances would you say that the people there gave orders, understood them, obeyed them, rebelled against them, and so on? Shared human behaviour [Die gemeinsame menschliche Handlungsweise[14]] is the system of reference by means of which we interpret an unknown language.
>
> (Wittgenstein PI, § 206)

> [The explorer in the foreign land] can come to understand it [the foreign language] only through its connections with the rest of the life of the natives. What we call 'instructions', for example, or 'orders', 'questions', 'answers', 'describing', etc. is all bound up with very specific human actions and an order is only distinguishable

> as an order by means of the circumstances preceding or following //accompanying it //.
>
> Wittgenstein (n.d. MS 165, 97F)

All of this includes both natural and nurtured behaviour. Moreover, even that which is culturally specific is ultimately rooted in biology (see § 1, above). Gordon Baker and Peter Hacker have made a similar point:

> Shared human behaviour provides the essential leverage for understanding mankind. This 'shared behaviour' is not only the common behaviour of mankind which manifests our animal nature, our natural needs for food, drink, warmth, our sexual drives, our physical vulnerability, etc. It also includes the culturally specific forms of behaviour shared by members of the tribe—their specific forms of social behaviour—observation of which and interaction with which enables us to interpret their language.
>
> (Baker and Hacker 2009, 173)[15]

In a footnote, Baker and Hacker add that 'any "form of life" accessible to lions, given their natural repertoire of behaviour and their behavioural dispositions, is too far removed from ours for any noises they might emit to count as speech' (Baker and Hacker 2009, 173 n. 1; cf. 1985, 186 n.1; see also 1985, 328ff.; 2009, 218ff.).

How does shared human behaviour aid us in interpreting an unknown language? By enabling us to establish regular connections between speech and action. This provides us with the leverage for interpreting others and, through this, to be in a position where we can be in receipt of reasons. In the absence of *any* such discernible regularity, 'we' cannot confidently say of the 'alien' people that they speak a language (Baker and Hacker 2009, 176; cf. 1985, 189–190). But even lions have sufficient regularity between their growls and their behaviour for one to at least take the crucial first step onto the ladder of understanding.

Von Savigny has pointed out that the pertinent behaviour need not be common to all (or even most) human beings (von Savigny 1991, esp. 113–117). What matters is that it is shared by both interpreter and interpretee. For this to be the case, its meaning must be public, in Wittgenstein's sense of the term; that is, that they can, in principle, come to learn it (see Baker and Hacker 2009, 177; cf. 1985, 190).[16] The worry, however, is not merely that feline noises do not count as speech but that even if lions did speak (in words or whatever), 'we' would not be able to understand them anymore than we could if we were to, *per impossibile*, have direct access to their minds (see Wittgenstein 2009, § 284; cf. Davidson 1984; Schroeder 2019; Sandis 2019b).

Shareability is a necessary but insufficient condition for understanding others. It is all too easy to systematically misinterpret behaviour,

linguistic or otherwise.[17] *Pari passu*, recognising that someone has a language is not *ipso facto* understanding that language. But Wittgenstein is not presenting us with a theory of understanding which specifies necessary *and* sufficient conditions for its existence, nor is he analysing the concept of understanding by breaking it down to its most basic constituents. Rather, he is probing for *limiting cases* of resemblances associated with understanding others.

The lion remark is not really about lions at all, let alone animals in general; it is about the very notion of what it takes to understand another. The barrier to understanding the speaking lion is neither metaphysical nor epistemic. It is—like forms of life themselves—partly biological and partly cultural. Wittgenstein holds that it is virtually impossible to understand another being not because of any *a priori* impossibility but because it can be extremely hard to do so biologically, culturally, and psychologically, depending on both the type and the individual.[18]

Epilogue

It begs the question to assume that if a lion could speak it would say things like 'gnaw me this cord' or 'fetch me an antelope'. For to know what sorts of things the lion would say if it could speak is to already understand it. It is, thus, far from obvious that we would understand a talking—or even a speaking—lion. Lions are hardly spherical Martians, and Wittgenstein is not claiming that it is logically or metaphysically impossible to understand them. Nor does he have any interest in whether it is empirically impossible for even Mowgli to do so. Perhaps lions and humans can both learn (or be trained) to display behaviour that is not just shareable but actually shared. By and large, however, this does not come to them naturally.

Human and many non-human animals alike typically share certain basic biological features (birth, sex, hunger, death). These pave the way for what Glock refers to as 'behavioural universals' (Glock 1996b, 166); that is, shared behaviour at a very general level (giving birth, eating, grieving, etc.). Such 'patterns of behaviour', in turn, presuppose 'a framework of shared cognitive capacities, needs, emotions, and attitudes' (Glock 1996b, 169). This is so even when the related rule-governed practices differ wildly from species to species and, indeed, within subgroups of each (cf. Vico 1725/1999, §§ 332–333; Sharpe 2005). Exclamations of hunger, gestures of disapproval, expressions of love, and practices of mourning for the dead may differ radically across different periods and places, but the general practices are found in both human and many non-human forms of life and it does not take much for outsiders to recognise them as such:

> It is for the most part straightforward to distinguish threatening and submissive gestures, since these are tied up with characteristic forms

> of human action, and the gesture of pointing is shared by all known [human] cultures.
>
> (Glock 1996b, 167)

While most non-human animals are notoriously bad at understanding how *human* pointing works, studies in animal gesture show that many species (e.g. goats, elephants, apes, and dogs) understand what it is to point.[19] More *to* the point, they can tell when we are being threatening and when we are being submissive and we can do the same for them.

I have never been on safari with Professor Glock, but I am sure that if we did go, we would both be able to distinguish a lion that was hungry from one that was full. The only trouble is that Hanjo can easily outrun me.[20]

Notes

1 Glock quotes from the original Anscombe translation, which was the only one published at the time. The translation as revised by Schulte and Hacker reads 'If a lion could talk, we wouldn't be able to understand it' (Wittgenstein 2009, § 327). I return to questions of modality further below and to the perils of translating 'sprechen' as 'talk'.
2 For this and more on the distinction between speaking and talking, see Sandis (2012).
3 As we shall see in § II, the question of who 'we' are is tricky for Wittgenstein and we should be suspicious of the thought that within the context of the lion remark, 'we' stands for *all* humans.
4 I am assuming (and assuming that Glock is assuming) that the growls and roars cannot be combined to express propositions. Whatever meaning they might carry is not conveyed in a word-like manner.
5 There is a parallel here between understanding what someone *says* and what they *write* or, indeed, *sculpt*. Understanding a work of art and understanding the artist are two different things (see Sandis 2017a, 2017b).
6 In an earlier draft of paragraph 327, the line about the lion is immediately followed by the statement 'He becomes an enigma, enigmatic to us through a certain kind of behaviour' (Wittgenstein n.d., MS 167, 12v–13r, as quoted in Schroeder 2019, § 4).
7 See, for example, the opening remarks in Wittgenstein (CV).
8 Wittgenstein would have read tales of talking animals by the Brothers Grimm and would also have no doubt wondered what exactly was being imagined in Carroll (1871), Kipling (1894), Baum (1900), Burroughs (1914), and Lewis (1950).
9 By contrast, Severin Schroeder treats the remark as an aphorism 'that is to be interpreted on its own without any context' (Schroeder 2019, § 4).
10 Both passages are quoted in full in §§ 1and 2 above.
11 See also Glock (1996b, p. 171 n. 28).
12 The complete list, as detailed by Hacker (2015), is Wittgenstein PI, §§ 19, 23, 241; n.d. MS 160, 51; 2009, §§ 1, 345.
13 Translation taken from Baker and Hacker (2009, 176, cf. 1985, 190, which also includes the original German).
14 See von Savigny (1991, 113–114) for an account of why Anscombe's rendition of this phrase as 'the common behaviour of mankind' is problematic.

15 In this revised version of Baker and Hacker (1985, 186–187), they have substituted 'shared behaviour' for 'common behaviour' throughout, thereby making it clearer that the common behaviour *of humanity* does not completely exhaust our shared behaviour and that the latter also includes behaviour that is 'culturally specific' (a term that helpfully replaces what was previously described as 'the diverse species-specific forms which such behaviour may naturally take for human beings').
16 This is compatible with something that is contingently hidden: 'If I were to talk to myself out loud in a language not understood by those present my thoughts would be hidden from them' (Wittgenstein 2009, § 317).
17 This might be as basic as a horizontal/vertical nod disparity or as complex as the structured misunderstandings in Jane Austen's *Pride and Prejudice*.
18 In keeping with this, Wittgenstein frequently complained that nobody understood him and/or his work. This was not vanity, for he just as often claimed that he was incapable of understanding other thinkers and/or their work (see Sandis 2015, 2016).
19 See, for example, Nawroth et al. (2020) and Smet and Byrne (2013, 2020).
20 Many thanks to Nicole Rathgeb and Eva Schmidt for their incredibly helpful comments on an earlier draft.

References

Anscombe, G.E.M. (1957). *Intention*. Oxford: Blackwell.

Baker, G.P. and Hacker, P.M.S. (1985). *Wittgenstein: Rules, Grammar and Necessity: Essays and Exegesis of §§ 185–242*. Vol. 2 of *An Analytical Commentary on Wittgenstein's 'Philosophical Investigations'*. Oxford: Blackwell.

Baker, G.P. and Hacker, P.M.S. (2009). *Wittgenstein: Rules, Grammar, and Necessity*. Rev. Oxford: Blackwell.

Baum, F. (1900). *The Wonderful Wizard of Oz*. Illustrated by W.W. Denslow. Chicago, IL: George M. Hill.

Burroughs, E.R. (1914). *Tarzan of the Apes*. Chicago, IL: A.C. McClurg.

Carroll, L. (1871). *Through the Looking Glass and What Alice Found There*. Illustrated by J. Tenniel. London: Macmillan.

Davidson, D. (1984). Thought and Talk. In *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, pp. 155–170.

Dennett, D.C. (1979). 'Intentional Systems'. In *Brainstorms: Philosophical Essays on Mind and Psychology*. Montgomery, VT: Bradford Books, pp. 3–22.

Dennett, D.C. (1991). *Consciousness Explained*. New York: Little, Brown and Co.

Dennett, D.C. (1995). Do Animals Have Beliefs? In Roitblat, H.L. and Meyer, J.-A. (eds.), *Comparative Approaches to Cognitive Science*. Cambridge: MIT Press, pp. 111–118.

Dennett, D.C. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.

Flaubert, G. (1913). *Le dictionnaire des idées reçues*. Paris: L. Conard.

Glendinning, S. (1998). *On Being with Others*. London: Routledge.

Glock, H.-J. (1996a). *A Wittgenstein Dictionary*. Oxford: Blackwell.

Glock, H.-J. (1996b). On Safari with Wittgenstein, Quine and Davidson. In Arrington, R. and Glock, H. (eds.), *Wittgenstein and Quine*. London: Routledge, pp. 144–173.

Glock, H.-J. (2000). Animals, Thoughts and Concepts. *Synthese* 123: 35–64.
Glock, H.-J. (2006). Thought, Language and Animals. In Kobe, M. (ed.), *Deepening Our Understanding of Wittgenstein*, special issue of *Grazer Philosophische Studien* 71: 139–160.
Hacker, P.M.S. (2015). Forms of Life. *Nordic Wittgenstein Review* 4: 1–20.
Jakobson, R. (1957/1984). Shifters, Verbal Categories and the Russian Verb. In Waugh, L.R. and Halle, M. (eds.), *Russian and Slavic Grammar: Studies 1931–1981*. Berlin: De Gruyter, pp. 41–59.
Kipling, R. (1894). *The Jungle Book*. London: Macmillan.
Kursell, J. (2010). First Person Plural: Roman Jakobson's Grammatical Fictions. *Studies in East European Thought* 62(2): 217–236.
Lewis, C.S. (1950). *The Lion, the Witch and the Wardrobe*. Illustrated by P. Baynes. London: Geoffrey Bles.
Nawroth, C., Martin, Z.M. and McElligott, A.G. (2020). Goats Follow Human Pointing Gestures in an Object Choice Task. *Frontiers in Psychology* 11(Art. 915): 1–6.
Pitcher, G. (1964). *The Philosophy of Wittgenstein*. Englewood Cliffs, NJ: Prentice-Hall.
Sandis, C. (2012). Understanding the Lion for Real. In Marques, A. and Venturinha, N. (eds.), *Knowledge, Language and Mind: Wittgenstein's Thought in Progress*. Berlin: de Gruyter, pp. 138–161.
Sandis, C. (2015). If Some People Looked Like Elephants and Others Like Cats: Wittgenstein on Understanding Others and Forms of Life. *Nordic Wittgenstein Review* 4: 131–153.
Sandis, C. (2016). Period and Place: Collingwood and Wittgenstein on Understanding Others. *Collingwood and British Idealism Studies* 22(1): 171–198.
Sandis, C. (2017a). If an Artwork Could Speak: Aesthetic Understanding After Wittgenstein. In Hagberg, G. (ed.), *Wittgenstein on Aesthetic Understanding*. Cham: Springer International.
Sandis, C. (2017b). The Doing & the Deed: Action in Normative Ethics. *Royal Institute of Philosophy Supplements* 80: 105–126.
Sandis, C. (2019a). Who Are "We" for Wittgenstein? In Appelqvist, H. (ed.), *Wittgenstein and the Limits of Language*. London: Routledge, pp. 172–196.
Sandis, C. (2019b). Making Ourselves Understood: Wittgenstein and Moral Epistemology. *Wittgenstein-Studien* 10(1): 242–260.
Schroeder, S. (2019). God, Lions, and Englishwomen. In Galvez, J.P. and Gaffal, M. (eds.), *Human Understanding as a Problem*. Berlin: De Gruyter, pp. 171–184.
Sharpe, L. (2005). *Creatures Like Us?* Exeter: Imprint Academic.
Smet, A.F. and Byrne, R.W. (2013). African Elephants Can Use Human Pointing Cues to Find Hidden Food. *Current Biology* 23(20): 2033–2037.
Smet, A.F. and Byrne, R.W. (2020). African Elephants Interpret a Trunk Gesture as a Clue to Direction of Interest. *Current Biology* 30(16): R926–R927.
Vico, G. (1725/1999). *New Science: Principles of the New Science Concerning the Common Nature of Nations*. 3rd revised edition. Translated by D. Marsh. London: Penguin.
Von Savigny, E. (1991). Common Behaviour of Many a Kind: *Philosophical Investigations* Section 206. In Arrington, R. and Glock, H.-J. (eds.), *Wittgenstein's Philosophical Investigations*. London: Routledge, pp. 105–119.

Wittgenstein, L. (1953). *Philosophical Investigations* [=PI]. Translated by G.E.M. Anscombe. Oxford: Blackwell.

Wittgenstein, L. (1967). *Zettel* [=Z]. Translated by G.E.M. Anscombe. Oxford: Blackwell.

Wittgenstein, L. (1998). *Culture and Value* [=CV]. 2nd revised edition. Revised by A. Pichler. Edited by G.H. von Wright. Translated by P. Winch. Oxford: Blackwell.

Wittgenstein, L. (2009). Philosophy of Psychology: A Fragment [=PPF]. In *Philosophical Investigations*. 4th edition. Revised by P.M.S. Hacker and J. Schulte. Oxford: Wiley Blackwell, pp. 182–264.

Wittgenstein, L. (n.d.). *Nachlass*. MS 160, 165, 167; TS 232. http://wab.uib.no/transform/wab.php. Last accessed: 28 March 2022.

Part II

Metaphilosophy, truth, and perception

5 The concept of truth

A proposal for a definition and its presuppositions[1]

Wolfgang Künne

In this chapter, I will argue (not for the first time) for a definition of the concept of truth that originates from a truism. Wittgenstein formulated this truism about truth in the 1930s when he wrote:

> ($W._1$) '*Was er sagt, ist wahr* = *Es verhält sich so, wie er sagt*.'
> (What he says is true = Things are as he says.)
> (PG, 123; cf. TLP 4.062)

Since a truth candidate need not be put into (perceptible) words, Wittgenstein would probably not mind the following revision of his formulation:

> ($W._1$)$^+$ What a person says or thinks is true if, and only if, things (really) are as she says or thinks they are.

I have replaced Wittgenstein's equality sign with the connective 'if, and only if', since the equality sign, just like the identity operator, cannot properly be placed between full sentences. The saying and thinking referred to in this truism is not always an asserting or a believing, for one can say something that is true without putting it forward as true, without *asserting* it, and one can entertain a true thought without acknowledging it as true, without *believing* it.

What is it that truth is ascribed to in Wittgenstein's dictum? What Hans asserted today in Hamburg, Jean had already asserted the day before yesterday in Paris, and John asserted it yesterday in London (each using his mother tongue to say it). Amie, a Scot, believes what the three gentlemen said while Betty, a Welshwoman, does not. What is the one thing stated by the three gentlemen, believed by one of the two ladies, and disbelieved by the other? That the first of February last year was a sad day for the United Kingdom and for Europe.[2] This 'one thing' is not a *sentence* of this or that language, let alone an *utterance* of a sentence of this or that language. Nor is it the linguistic *meaning* of a sentence, for in earlier years one would have said something different (different even in truth-value) by uttering the sentence 'The first of February *last* year was a sad day for the UK and for Europe' or a correct translation of

DOI: 10.4324/9781003202929-8

it, although the linguistic meaning of the sentence was the same then as it is now. The 'one thing' that various people put forward as true, take to be true, or take to be false is obviously not a *mental act* or *state* either. Following a usage that seems to have originated in Cambridge early in the twentieth century, I call this 'one thing' a *proposition*.[3]

When one asserts that things are as John says or thinks they are, one does not ascribe a particular structure to what he said or thought. Wittgenstein's truism does not bring with it a commitment to the idea that truths can always be expressed by sentences that are used to ascribe a property to one or more objects or that every truth is expressible in a language with the structure of the classical predicate calculus. This reticence is well-advised, for explanations of the concept of truth which, like those of Bolzano or Tarski, are committed to the first or to the second form requirements respectively, are too narrow unless every candidate for the title 'true' has the required structure. It speaks in favour of an explanation of the concept of truth if it does not have to force any candidate for truth onto a Procrustean bed of certain formal requirements in order to integrate it. ('And if thou'rt unwilling, then force I'll employ'—the force of imposing paraphrases on sentences, the appropriateness of which may be, and always remain, debatable.)

In our truism, the use of 'as' deserves particular attention. It is the starting point of all correspondence theories of truth, and it is the starting point—and almost the endpoint, too—of the definition I propose. The Greek and Latin counterparts are 'οὕτως ὡς' and 'ita sicut'. In Plato, the sophist Hippias concedes to his interlocutor: '*ἔστι μὲν ταῦτα, ὦ Σώκρατες, οὕτως ὡς σὺ λέγεις* (Things are as you say, Socrates)'. William of Ockham writes: '*propositionem esse veram est ita esse sicut significatur per propositionem* (the being true of a sentence consists in things being as the sentence says)' (Plato Hp.Ma. 282a4; William of Ockham OPh II, 377). Strawson uses the expression 'as' when he formulates our truism for two special cases:

> (S) 'A statement is true if and only if things are *as* one who makes that statement thereby states them to be. A belief is true if and only if things are *as* one who holds that belief thereby holds them to be'.
> (1976, 273, my emphasis)

How are these comparative particles to be understood? Let us consider the following intuitively valid argument:

> (P1) Galileo says that the earth moves, and the earth (really) moves. Therefore,
> (C1) at least sometimes things (really) are as Galileo says they are.

The word 'really' is here and hereafter only a reading aid—it is semantically redundant. In premise (P1), we express the point of the 'as' in the

conclusion by means of a *conjunction* in which the sentence 'the earth moves' serves two purposes: In the first conjunct, it is used to specify the content of a linguistic utterance; in the second conjunct, it is used to make a cosmological statement. Let us try to express the conclusion as a conjunction as well:

> (C1)′ Sometimes Galileo says that things are thus and so,
> and this is how things (really) are.

To express this more concisely, we can avail ourselves of another of Wittgenstein's observations:

> (W.$_2$) '*Es verhält sich so und so*' [. . .] *ist* [. . .] *ein deutscher Satz, denn es hat Subjekt und Prädikat. Wie aber wird dieser Satz angewendet* [. . .]? [. . .] *Er wird als Satz<u>schema</u> verwendet.* [. . .] *Man könnte auch, wie in der symbolischen Logik, bloß einen Buchstaben, eine Variable gebrauchen.* [. . .] [O]*bschon es ein Satz ist, so hat es doch nur als Satzvariable Verwendung.* ('This is how things are' is an English sentence, for it has a subject and a predicate. But how is this sentence applied? It is employed as a propositional *schema*. One could instead simply use a letter, a variable, as in symbolic logic. Though it is a sentence, still it gets used only as a propositional variable.)
>
> (PI, § 134)

The ascription of a subject-predicate structure to the English sentence referred to in the translation is rather dubious, and it is definitely incorrect as regards the German sentence. The pronoun '*es*' does not serve here as a subject (term), for the question whom or what it designates in an utterance of this sentence does not arise here any more than in the case of 'It' in 'It is raining'. Compare the use of '*es verhält sich so*' in

> (S1) *Das Mädchen (der Junge) hat heute getobt und um sich geschlagen,*
> *und das ist nicht der erste Tag, an dem es (er) sich so verhält.*

In the second conjunct, '*es verhält sich so*' is not a placeholder for a whole sentence. Here, the word '*es*' really is a subject term that inherits its reference from its antecedent '*das Mädchen*', and '*verhält sich so*' really is a grammatical predicate that functions like a predicate schema, like a placeholder for a predicate.

With the help of bound sentence variables, we can formulate the conclusion of the argument '(P1), therefore (C1)' as follows[4]:

> (C1)″ $\exists p$ ((what Galileo says is that p) & p).

If we now replace the phrase 'what Galileo says' with a variable whose values are propositions, we get the definiens of the definition of 'x is true' that I advocated under the name 'A Modest Account of Truth' in my *Conceptions of Truth*[5]:

(Df. T) x is true: $\leftrightarrow \exists p$ ((x is the proposition that p) & p)

('Quantificational' or even 'Quantificational-Conjunctive', though deterrently technical, would have been more telling than 'Modest', and these adjectives would not have carried the unpleasant suggestion that the proponent of (Df. T) takes his opponents to suffer from the vice of immodesty. But within the book, the point of calling the account 'modest' was reasonably clear.)

Like all other attempts at defining the concept of truth, (Df. T) has come under fire. In the following pages, I will discuss three questions that critics have raised and that I deem to be especially challenging.[6]

Problem No. 1, or How to make sense of the *sentential quantifier* in (Df. T)

The fact that the explanation of truth I propose makes use of quantification into the position of a full sentence raises the question whether it is possible to make such quantification intelligible without drawing on the concept of truth. If it is not possible, then the explanation is circular. This suspicion of circularity could not be dismissed if the quantifier in (Df. T) had to be understood as a *substitutional* quantifier, for then (Df. T) would tell us that something is true just in case at least one substitution instance of the open sentence after the quantifier is true (expresses a truth). Thus understood, not only would (Df. T) be worthless,—we would also not be entitled to maintain what I consider to be absolutely correct (I will come back to this below):

($\exists$1) $\exists p$ (the thought that p will never be entertained, let alone be put into words),

for in a substitution instance of the open sentence after the quantifier a thought *is* put into words in the 'p' position, and of that thought the substitution instance goes on to say that it will *never* be expressed. The sentential quantifier I use in (Df. T) is just as non-substitutional as the quantifier '$\exists x$' under its standard objectual reading. Under this reading, '$\exists x$' allows for a sentence like ($\exists$2) to express a truth:

($\exists$2) $\exists x$ (x is never individually referred to).

Of the flowers in Thomas Gray's *Elegy Written in a Country Churchyard*, it is probably true that none of them is ever pointed to or otherwise individually referred to:

> *Full many a flower is born to blush unseen,*
> *And waste its sweetness on the desert air.*

And this certainly holds of countless *grains of sand* in the Sahara.

Those philosophers who frown at the sentential quantifier in (Df. T), if it is to be understood non-substitutionally, should first of all be reminded—propaedeutically, as it were—that there are more quantifiers than are employed in first-order classical logic. The premise

> (P2) Today Ann is busy in her office in London

does not only entail quantifications into the position of singular terms, for example,

> (C^{sgl}) Today *somebody* is busy in Ann's office in London

but also

> (C^{gen}) There is something that Ann is today (namely, *busy in her office in London*),

where (as the 'namely' rider shows) we quantify into the position of a general term. In fact, no logician should be uncomfortable with this inference. After all, Frege, our venerable *quantifex maximus*, also makes use of quantification into the position of predicates in his *Begriffsschrift*, and quantification into the position of general terms is the colloquial counterpart to this. Second, (P2) entails:

> (C^{temp}) *At least once* Ann is busy in her office in London (namely, *today*),

where (as the rider shows) we quantify into the position of a temporal adverb. Third, (P2) entails:

> (C^{loc}) Today Ann is busy *somewhere* (namely, *in her office in London*),

where (as the rider shows) we quantify into the position of a local adverbial.—Granted, the critic might concede (perhaps grudgingly), in colloquial language we quantify not only into the position of singular terms but also into that of general terms and adverbials, but do we also quantify into the position of full sentences?

I think the (non-substitutional) sentential quantifier in (Df. T), which is not to be understood substitutionally, can be made intelligible in two steps. *First step*. Compare (S2) with (S3):

(S2) She loves *Mozart's first piano quartet*, and he loves *it* too.

(S3) According to her report, *the President and his son are involved in a crime*. If *this is how things are* the President should be removed from office.

One can recognise a revealing *analogia proportionalitatis* here. The 'this is how things are' in (S3) stands in the same relation to the italicised sentence in (S3) as the 'it' in (S2) stands to the italicised nominal phrase in (S2). The 'This is how things are' in (S3) owes its content to its antecedent, the sentence for which it deputises; it is an anaphoric prosentence,—just as the anaphoric pro*noun* in (S2) gets its content from its antecedent, the nominal phrase for which it deputises. Both expressions are pro-forms that save us from having to repeat ourselves, 'proforms of laziness'. The prosentence in (S3) saves us from having to repeat a sentence, as do the laconic anaphoric prosentences 'yes' and 'so' in

(S4) Is the manuscript too long?—Yes.
(S5) Is the manuscript too long? If so, you should shorten it.

(Franz Brentano introduced the term '*Fürsatz*', as opposed to '*Fürwort*', to characterise the role of the little word '*ja*' that occurs in the German translations of (S4) and (S5). Unfortunately, many German analytic philosophers only know the American counterpart of Brentano's term, 'prosentence', which is why there is a philosophical literature on *Prosentenzen*, as if any old sentence were a *Sentenz*.)

Second step. Compare the following arguments:

[A1] Generally, if something is a cube then it has eight corners.
Therefore,
if the Kaaba is a cube then the Kaaba has eight corners.

[A2] Generally, if Ann says that things are thus and so then this is how things are.
Therefore,
if Ann says that the President is a liar then the President is a liar.

In both arguments, we draw an inference from the general to the particular. Here, another *analogia proportionalitatis* catches the eye. The occurrence of 'this is how things are' in the premise of [A2] relates to its occurrence in (S3) in the same way as the occurrence of 'it' in the premise of [A1] relates to its occurrence in (S2). What in (S2) and (S3) was

an anaphoric proform is a quantificational proform in [A1] and [A2]—colloquial counterparts of a bound variable in a formal language. [A1] is the colloquial equivalent of

[A1]* $\forall x$ (x is a cube $\rightarrow$ x has eight corners). Therefore . . .

and [A2] is the colloquial equivalent of

[A2]* $\forall p$ ((Ann says that p) $\rightarrow$ p). Therefore . . .

For all I can see, the preceding explanation of the sentential quantifier did not at any point draw on the concept of truth.—You will have noticed that, according to (Df. T), the premise in [A2] or [A2]* expresses the thought that everything Ann says is *true,* and it does so without using the word 'true' or a synonym.

After having explained quantification into the position of full sentences in this way, one can use it in the metalanguage of a semantic theory that specifies the conditions for truth-under-an-interpretation for formulae containing sentential quantifiers (cf. Williamson 1999, 263, 2003; esp. Rumfitt 2011, 18–9). The same thing happens in the metalanguage of a semantic theory for formulae containing nominal quantifiers. In the years following the publication of *Conceptions of Truth*, I have ruefully returned to the view of sentential quantification that I held in *Abstrakte Gegenstände*: It is a *sui generis* form of quantification in which nothing is quantified over.

Problem No. 2, or How to elucidate the term '*proposition*' in (Df. T)

Can the meaning of the philosophical term 'proposition' be explained without appeal to the concept of truth? If not, (Df. T) is circular. I think the following explanation of the meaning of 'proposition' as used in the definiens of (Df. T) is not open to the charge of circularity.

The explanation proceeds in three steps. *First step*. Whenever a clause of the form 'that *p*' is used to specify what somebody thought or said (in whatever psychological or illocutionary mode), the term 'the proposition that *p*' designates the content of this thinking or saying. The general term 'proposition' is the least specific of a whole battery of commonly used general terms that can fill the slot in 'the [. . .] that *p*'. Depending on the psychological or illocutionary mode of the thinking or saying, or on the role its content plays e.g. in a conversation, in a theory, or in a religion, the slot may be filled by 'answer', 'assertion', 'assumption', 'axiom', 'belief', 'claim', 'conviction', 'doctrine', 'dogma', 'hypothesis', 'law', 'opinion', 'principle', 'statement', 'supposition', 'theorem', etc. (Perhaps this observation helps to demystify propositions.)

Second step. 'Content' was the key term used in Step One. How is it to be understood? Let me elucidate it by exemplification, using belief and assertion as my examples. If Ben's belief that one day the dead will rise is unshakeable, while Ann's belief that one day the dead will rise is easily shaken, it does not follow that one and the same entity is at the same time unshakeable and not unshakeable. In both cases, the term 'belief' applies to a mental state, the identity and existence of which stand and fall with the identity and existence of its bearer. But in another sense of 'belief', Ben and Ann *share* for a while a certain eschatological belief. In other words, his unshakeable belief has the same *content* as her easily shaken belief. The same applies *mutatis mutandis* to assertions. When we say 'In New York, the audience responded with thunderous applause to Ann's assertion that the President is a notorious liar, while Ben's assertion in Miami that this is how things are provoked angry howls of protest', we are talking about assertions as datable illocutionary acts, the identity and existence of which stand and fall with the identity and existence of the relevant agents. But in another sense of 'assertion', Ann and Ben made the *same* assertion. In other words, her speech act had the same *content* as his.

Third step. The realm of propositions comprises not only propositions that are at some time or other the content of a saying or a thinking. Many of the things that have been said or thought would probably have better been left unsaid and unthought. But we can be sure that not everything which *could* be thought or said (in whatever psychological or illocutionary mode) will ever *actually* be thought or said. This includes, for example, all answers to the maximally uninteresting question how often the letter A occurred last year on the pages of the *Daily Mirror*. Presumably, this question was formulated here for the first time, and nobody will ever bother to answer it. So countless propositions that could be expressed by substitution instances of the open sentence 'Last year the letter A occurred *n* times on the pages of the *Daily Mirror*' will never be the content of any thinking or saying. Many thinkables and sayables remain forever unthought and unsaid. (It is not advisable to comply with the request to give an example of something that could be thought or said without actually ever being thought or said.)—This reminder completes my attempt at explaining the concept of a proposition as used in (Df. T). For all I can see, there was no appeal to the concept of truth at any point.

Of course, I agree with the claim that it is a necessary a priori truth that propositions are 'truth-evaluable', that is, that they are things which can be evaluated in the dimension of truth and falsity. However, this does not entail that one cannot explain the meaning of the term 'proposition' without recourse to the concept of truth. It is also a necessary a priori truth that equilateral triangles are equiangular; but that does not entail that one cannot explain the meaning of the phrase 'equilateral triangle' without recourse to the concept of equiangularity.

By maintaining that not all sayables and thinkables will ever actually be said or thought, one does not incur the metaphysical commitment (as one of my critics presumes (Boghossian 2010, 563)) that propositions are 'mind-independent' and 'language-independent' entities. That is to say, it does not put one under the obligation to accept the thesis

> (Ind_1) $\forall p$ (the proposition that p exists even in a world in which nothing is ever thought or said).

The independence claims that are actually entailed by my conception of propositions are much weaker:

> (Ind_2) $\forall p$ (the proposition that p exists even in a world in which nobody ever says or thinks that p)
> (Ind_3) $\forall p$ (the existence of the proposition that p does not depend on anyone's ever saying or thinking that p).

(Ind_2) and (Ind_3) can be true even if (Ind_1) is false. A comparison may be helpful. Let us look at Uncle Tom's cabin. There is no log of which one could truly maintain: If *this* log did not exist, then Uncle Tom's cabin would not exist either. But a world without any logs would be a world without Uncle Tom's cabin. Correspondingly, there is no saying or thinking that p of which one could rightly assert: If *this* act or state did not exist, then the proposition that p would not exist either. This is compatible with the claim that a world without any speakers or thinkers would be a world without propositions. *Individual* independence does not entail *generic* independence.

Problem No. 3, or How to defend the penchant for the truth *predicate* in (Df. T)

The expression which (Df. T) is supposed to define is a *predicate* in the Fregean sense of this word—an operator which, for singular terms as input, yields full declarative sentences as output. This predicate is clearly contained in the following three sentences:

> (T1) *What the author is trying to prove here* is true.
> (T2) *Goldbach's conjecture* is true.
> (T3) *The proposition that every even number greater than 2 can be represented as the sum of two prime numbers* is true.

The italicised parts of these three sentences are singular terms. They can go on each side of an identity operator: 'What the author is trying to prove here *is identical to* Goldbach's conjecture', 'Goldbach's conjecture

is nothing other than the proposition that every even number greater than 2 can be represented as the sum of two prime numbers'. Sentences in which 'true' is used as in (T1), (T2), and (T3) are within the scope of the definiendum of (Df. T).

Let us note in passing an important difference between the predications (T1) and (T2) on the one hand and (T3) on the other: In (T1) and (T2), the proposition to which truth is ascribed is not expressed but only designated. By contrast, in (T3) it is both designated and expressed.

So far, there is no scope problem in sight for (Df. T). This changes as soon as we consider the use of 'true' in sentences such as

(T4) It is true that snow is white.

This sentence can be parsed in the same way as logicians habitually parse the next two sentences:

(Neg) It is not the case that / snow is blue.
(Mod) It is possible that / snow is blue.

Logicians have discovered operators in (Neg) and (Mod) that yield sentences as output, not for singular terms, but for *full sentences* as input, and they have introduced symbols for these operators that are syntactically and semantically atomic. They call them (slightly paradoxically) *unary connectives*:

(Neg)* $\neg$ (snow is blue), ~ (snow is blue)
(Mod)* $\Diamond$ (snow is blue)

I prefer to call a connective that does not connect a sentence prefix or a *prefix* for short. Undoubtedly, one can also parse (T4) in such a way that a prefix comes to the fore:

(T4)′ It is true that / snow is white.

And we could (as Georg Henrik von Wright does in the calculi he calls 'truth-logics'[7]) introduce a symbol for the truth prefix that is syntactically and semantically just as seamless as the prefixes in propositional and modal logic:

(T4)* T (snow is white)

In this sentence, the proposition that snow is white is only *expressed* but not *designated*. Just as in (Neg)* and (Mod)*, the only thing that is designated in (T4)* is frozen water vapour. Arthur Prior, the pioneer of tense-logic, holds that (T4) *must be* analysed in such a way that the

sentence has the same structure as (T4)*, and he combines this claim with a priority thesis:

> (**PPT**) '[T]he word "true" in its primary use is an inseparable part of the adverbial phrase "it is true that"'.
>
> (1967, 229)

The expression that is mislabelled an adverbial phrase in Prior's Priority Thesis is a sentence prefix. The prefix is not really ad-verbial but rather ad-sentential. (PPT) implies that the use of the adjective 'true' as a component of the truth predicate is *not* its primary use.[8] If the truth prefix is indivisible, then nothing is classified *as true* in (T4): Only snow is classified as white. And if the use of 'true' as an inseparable part of the prefix is basic, as Prior claims, then an explanation of the concept of truth that is not primarily concerned with the prefix is a performance of *Hamlet* without the Prince of Denmark.[9] You will not be very surprised when I say: I am convinced that it is not the prefix but the predicate that deserves primacy.

According to (PPT), the truth prefix in (T4) is just as atomic as the calculus sign in (T4)*. This is patently false if it is to mean that the prefix in the English sentence (T4) is syntactically seamless ('It-is-true-that').[10] It has a seam between 'true' and 'that' where a parenthesis can be inserted and a complete that-clause can be cut off:

> (T4[a]) It is true, as we all know, that snow is white.
> (T4[b]) Alfred said that snow is white, and it is true [✂ that snow is white].

Unlike (T4[b]), 'Alfred said that snow is white, and it is true that' is not a grammatically well-formed sentence. Both observations show that the word 'that' is not chained to 'true'. So, the clause 'that snow is white' is a genuine syntactical constituent of (T4). This observation is not relevant if (PPT) attributes only *semantic* simplicity to the prefix in (T4). (After all, syntactic complexity does not exclude semantic simplicity. Consider 'Yesterday, Mr. X kicked the bucket' as the first line of a rather disrespectful obituary: Here, the syntactically complex predicate is semantically as simple as the synonymous predicate 'died'.) The crucial question is whether the meaning of the truth prefix contains the meaning of the truth predicate. Prior and his allies deny this, whereas I maintain it.

Let us take a closer look at (T4). The first word is a *pleonastic* or *expletive* pronoun: It is semantically vacuous; it does not contribute anything to the content of an utterance of (T4). If you look at the translation of (T4) into Italian, you will see that Italians do not allow themselves the luxury of an expletive: '*È vero che la neve è bianca*'. The first word

in (T4) has no more semantic content than the first word in Schiller's *Wilhelm Tell* where the fisherman's boy sings in his boat:

> (Tell) *Es lächelt der See, er ladet zum Bade.* (The lake smiles, it invites you to bathe)

What are the pronouns doing here? The '*er*' in the second sentence is an old acquaintance as regards its function: It is anaphoric and designates in the mouth of the fisherman's boy what its antecedent designates in the first sentence, that is, Lake Lucerne. As opposed to this, the '*es*' in the first sentence is a *pleonastic* pronoun: It does not denote anything, it is a semantic caecum.[11] If we are not concerned with the rhythm but merely with the propositional content, we can throw off the baggage and express the same proposition after a shift:

> (Tell)^0 *Der See lächelt, . . .*

We can do exactly the same in the case of (T4), which gives us

> (T5) That snow is white is true.

This sentence expresses the same proposition as (T4). Thus, if in utterances of (T5) we do not only classify snow as white but also a proposition as true then the prefix in (T4) *does* contain—contra Prior's Priority Thesis—the meaning of the truth predicate, and a definition of this predicate does not have to be supplemented by, let alone be based upon, an explanation of the prefix. It is exactly the other way round: The explanation of the prefix is based upon that of the predicate. But as I said, this is how things are *if* (T5) is a predication. After all, Prior can point out that one can also parse (T5) in such a way that an operator comes to the fore which yields sentences as output for sentences as input:

> (T5)$''$ That / snow is white / is true.

The only difference from (T4) is that the sentence-forming operator on sentences that is contained in (T5) is not the truth *prefix* 'It is true that ( )' but the truth *frame* 'That ( ) is true'. So (PPT) is back in the running, now in an enlarged version:

> (**PPT**)$^+$ The word 'true' in its primary use is a semantically inseparable part either of the truth prefix or of the truth frame.

In order to refute this primacy claim, we need to show that not only in utterances of (T1), (T2), or (T3) but also in the course of uttering (T4) or

(T5) the speaker singles out something in order to ascribe truth to it. So it needs to be shown that in utterances of (T4) or (T5) the speaker uses a *that-clause* for this job.

Before I try to show that, let me try to shield off a possible misunderstanding of my position. I do not deny that Prior's decomposition of (T4) and (T5) is permissible—I deny that it is *obligatory*. Everybody should concede that his parsing is *permissible*. After all, it pays tribute to the fact that every occurrence of (T4) or (T5) contains a *genuine* occurrence of the sentence 'snow is white'. An occurrence of this sequence of words within an English sentence is not always a genuine occurrence of the *sentence* composed of these words. Consider 'Not everybody who is fond of snow is white. (Many black people also like snow.)' In this environment, the meaning of the sentence 'snow is white' is not a component of the meaning of the compound sentence. But clearly, it is a component of the meaning of (T4) and (T5). So, in my view, these sentences really are substitution instances of the Priorese schemata 'It is true that *p*' and 'That *p* is true' respectively, but they are also instances of the predicational schema '*a* is true' where '*a*' is a placeholder for expressions of any complexity which are used to single out a truth-value bearer. As opposed to Prior, I regard recognition of the predicational form to be fundamental for our understanding of truth discourse.

The assumption that one and the same sentence cannot have more than one logical form seems to be due to false analogies. No diagram on the blackboard can be both square and round, and no sonnet can have both Petrarchan and Shakespearean form—true enough. But the sentence 'Every man is mortal' is a substitution instance of 'Every S is mortal', 'Every man is P', and 'Every S is P', so it has each of these forms. If you want to show which expressions are such that the validity of the hackneyed argument 'Every man is mortal, Socrates is a man, so Socrates is mortal' depends on their meaning, then only the Aristotelian assignment of the third form will do. But if you want to show which expressions are such that the validity of 'Every man is mortal, so every man is bound to die' depends on their meaning, then only the assignment of the first form is revealing, for 'Every S is mortal, so every S is bound to die' is the schema all substitution instances of which are valid. (Obviously, countless arguments of the form 'Every S is P, so every S is bound to die' move from a truth to a falsehood.)

It is agreed on all sides that the verbose subject term in (T3) serves to single out something to which truth can be ascribed. What needs to be argued for is the contention that the *naked* that-clauses in (T4) and (T5) play the same role. (I call a that-clause naked if it is not preceded by an apposition like 'the axiom', 'the belief', 'the claim', 'the doctrine', 'the hypothesis', 'the statement', or, as in (T3), 'the proposition'.)

In some contexts, at least, it seems undeniable that a naked that-clause serves this purpose. Consider the following pair of sentences:

(S6)[a] Platonism entails the negation of nominalism.
(S6)[b] That universals exist entails that not everything is a particular.

Entailment is a relation between truth-value bearers. In (S6)[a] the first relatum is singled out by a proper name and the second relatum by a definite description containing a proper name. (Only a few truth-value bearers have made it so far as to be given a proper name.) If we mean by 'platonism' the doctrine that universals exist and by 'nominalism' the doctrine that everything is a particular, then the same two truth-value bearers are singled out by the two naked that-clauses in (S6)[b]. Now consider

(S7) That snow is white entails that something is white, and it is true.

In the context of (S7), 'it is true' comes to the same thing as 'It is true that snow is white', for it is the result of cutting off. (Recall the scissor's work in (T4[b]).) If the first that-clause in (S7) is used to single out a truth-value bearer, so is the that-clause in (T4). Since we have seen reason to affirm the antecedent of this conditional, we should apply Modus Ponens. If the that-clause in (T4) serves to single out a proposition, then the clause in (T5) does so as well, for these sentences are just stylistic variants expressing the same proposition. (Recall '*Es lächelt der See*' / '*Der See lächelt*'.)

My argument nowhere depends on the controversial claim that all that-clauses are singular terms. If 'that *p*' is a singular term, then it designates what the singular term 'the proposition that *p*' designates, and one would expect that the naked that-clause and this singular term are interchangeable *salva veritate*, hence a fortiori interchangeable *salva grammaticalitate*. But 'She fears that *p*' does certainly not have the same truth-value as 'She fears the proposition that *p*',[12] and the result of replacing 'that *p*' in 'She is convinced that *p*' by 'the proposition that *p*' is ungrammatical. But in (S6)[b] the truth-value remains unaffected if we put 'the proposition' in front of both that-clauses, and the same holds for (T5), so why not say that *in such cases* the that-clause is a singular term?

Admittedly, in (T4) the that-clause does not tolerate this apposition. But how significant is this syntactical fact semantically? Sometimes an exchange of clearly co-referential terms does not preserve grammaticality. You see this when you replace 'Goethe' by 'the author of *Werther*' in

(S8) Young Goethe fell in love with Friederike.

But this grammatical fact is semantically not significant, for (S8) has the same sense as

(S9) When he was young Goethe fell in love with Friederike,

and here 'Goethe' *is* replaceable by 'the author Werther'. So, the lack of interchangeability in (S8) is just a semantically insignificant caprice of the grammar of the vernacular: Both in (S8) and in (S9), 'Goethe' is used to refer to the author of *Werther*. Now, (T4) is related to (T5) as (S8) is related to (S9). Since (T4) has the same sense as (T5), not only in (T5) but also in (T4) the that-clause is used to refer to the proposition that snow is white.

Here is my final reason for assigning primacy to the truth predicate. Consider the following argument:

[A3] (P_1) It is universally accepted that snow is white.
(P_2) It is true that snow is white.
Therefore,
(C) there is something which is both universally accepted and true.

It is agreed on all sides that this argument is valid: If the premises are both true, the conclusion cannot fail to be true. But why is this argument valid?

The argument 'Kant was a bachelor, therefore Kant was unmarried' is also valid. Why? Because of the recurrence of the same singular term and the meaning of the general terms 'bachelor' and 'unmarried'. These terms are not topic-neutral—they partly determine the topic of the arguments they appear in. Since Prior and his allies do not regard that-clauses as genuine semantical units, they have to put the argument [A3] into the same boat as the 'bachelor' argument. According to them, [A3] instantiates the following pattern:

f(P_1) It is universally accepted that / p.
f(P_2) It is true that / p.
Therefore,
f(C) $\exists x$ (x is true & x is universally accepted).

The argument is valid because of the recurrence of the same sentence, the meaning of two topic-neutral expressions, sc. the objectual existential quantifier and the connective '&', and the meaning of no less than four (non-synonymous) expressions, two of which are not topic-neutral. The truth prefix in (P_1) and the truth predicate in (C) differ in meaning, because two expressions that belong to different syntactical categories are never synonymous, and the same holds for the prefix 'It is universally accepted that' in (P_2) and the predicate 'is universally accepted' in (C). The acceptance prefix and the acceptance predicate are no more topic-neutral than the predicate 'bachelor' in the argument about Kant. Under the Priorese reading, [A3] is just as little *formally* valid as the 'bachelor' argument.

But does the validity of [A3] really depend on the meaning of the two prefixes and the two predicates? There is a far better explanation of the validity of [A3] if one regards that-clauses as genuine semantical units, for then its validity is entirely independent of the meaning of any expression that is not topic-neutral. Under this assumption, (P_1) and (P_2) are predications that ascribe the property of being universally accepted and the property of being true to one and the same proposition. If we move *from* two premises in which two different properties are ascribed to one and the same object *to* the conclusion that there is at least one object that has these two properties, we will never move from two truths to a falsehood, no matter what these properties are. Thus understood, [A3] has the form

$f(P_1)$ Fa
$f(P_2)$ Ga
Therefore,
f(C) $\exists x$ (Fx & Gx).

The argument is valid because of the recurrence of the same singular term and the meanings of two topic-neutral expressions, full stop. The meaning of the predicates in the positions of 'F' and 'G' is entirely irrelevant. In other words, the argument [A3] turns out to be *formally* valid.

For all these reasons, I endorse the anti-thesis to Prior's Priority Thesis. The word 'true' in its primary use is a component of the predicate '*x* is true'.[13]

Summary

I first tried to explain the structure of a *definition* of the concept of truth that sticks as closely as possible to a truism about truth. I then discussed three *presuppositions* of my proposal for a definition. It presupposes that my affirmative answers to the following three questions are correct: Can quantification into the position of a full sentence be explained without recourse to the notion of truth? Can the notion of a proposition be elucidated without invoking the notion of truth? Is it methodologically correct to give priority to the predicative use of 'true' in an explanation of the concept of truth? I have tried to justify my answers.

Notes

1 Many thanks to Nicole Rathgeb who provided me with a careful translation of my German text and with useful comments.
2 Written in September 2021.
3 I might as well have picked up a noun in Jena, '*Gedanke*'. The 'one thing' that in my scenario was put forward as true, taken to be true, and taken to be false was the *thought* that the first of February last year etc.

4 Cf. the use of the quantifier and the name variable in formalising (a) 'There is *someone* behind the tree and *he* is armed' as (b) '$\exists x$ (x is behind the tree & x is armed)'.
5 Various parts of what was to become that book received a thorough discussion in a seminar organised by my friend Hanjo Glock during my time as Leverhulme Visiting Professor at Reading University.
6 The criticisms I found, and still find, most instructive are those by Paul Boghossian (2010), Kevin Mulligan (2010), and Ian Rumfitt (2011).
7 In von Wright's truth calculi, exactly one symbol is added to the vocabulary of classical propositional logic, viz. the prefix 'T', and we are asked to read it as 'it is true that'. Thanks to the introduction of 'T', it is possible to do something in this extended calculus that is not possible in classical propositional logic: One can admit the possibility of truth-value gaps within the calculus, that is, distinguish the non-truth of a truth candidate from its falsity. The lack of truth is expressed by '$\neg$ T p' and falsity by 'T $\neg$ p'. Cf. von Wright 1984, 26–41; 1986; 1996, 71–91 and the discussion in Morscher 2015.
8 This is also the main contention in Mulligan 2010. For a very extensive discussion of his position (including his substitution worries that I have put aside in this paper), see Künne 2014.
9 Due to the so-called '*Regietheater*', this popular metaphor is in danger of becoming obsolete in Germany.
10 For all I know, this was first shown in Wilson 1990, 23–24.
11 German Lieder and poems abound with this construction: '*Es ist ein Ros' entsprungen*' (Christmas hymn), '*Es schlug mein Herz, geschwind zu Pferde!*' (Goethe), '*Es bellen die Hunde, es rasseln die Ketten*' (Müller/Schubert), '*Es klappert die Mühle am rauschenden Bach*' (folk song), . . . I sense a similarity with the pronoun in the Irish Regiment's song 'It's a long way to Tipperary', which seems to come to the same thing as 'The way to Tipperary is long'.
12 Even neurotic nominalists will not fear any proposition, for they believe that there are no propositions.
13 The truth predicate, like any other predicate of the form ' . . . is (an) F', is not unbreakable either. The arguments in Wilson 1990 can be easily adapted to show that general terms are genuine constituents of predicates. Parenthesis insertion is possible: 'That contention is, as you should know, true', and so is cutting off: 'Not every statement in his book is true, but this one is [✂ true]'. Further, in quantifications like 'She is something he will never be', we quantify into the position of a general term, as the survival of the copula 'is' and the rider 'namely witty' shows. So, Quine is wrong when he says that a predicate like ' . . . is true' and ' . . . is witty' is 'an indissoluble unit in which [the general term] stands merely as a constituent syllable' comparable to the 'tic' in 'attic' (Quine 1974, 220). Not even a verb-only predicate like ' . . . smiles' is unbreakable, since we can quantify into the position of the verb-stem: 'She often smiles, but he never does. So, she often does something he never does, namely smile'. Cp. the homage to the copula in my 2006, §§ I and V.

References

Boghossian, P. (2010). Our Grasp of the Concept of Truth: Reflections on Künne. *Dialectica* 64(4): 553–563.

Künne, W. (1983/2007). *Abstrakte Gegenstände*. 2nd edition. Frankfurt am Main: Klostermann.

Künne, W. (2003). *Conceptions of Truth*. Oxford: Clarendon Press.

Künne, W. (2006). Properties in Abundance. In Strawson, P.F. and Chakrabarti, A. (eds.), *Universals, Concepts and Qualities: New Essays on the Meaning of Predicates*. Aldershot: Ashgate, pp. 249–300.

Künne, W. (2014). Truth without Truths? 'Propositional Attitudes' without Propositions? Meaning without Meanings? In Mulligan, K. et al. (eds.), *The History and Philosophy of Polish Logic*. Houndsville, AL: Palgrave, pp. 160–204.

Morscher, E. (2015). The Logic of Truth. *Grazer Philosophische Studien* 91: 409–421.

Mulligan, K. (2010). The Truth Predicate vs the Truth Connective: On Taking Connectives Seriously. *Dialectica* 64(4): 565–584.

Plato. Hippias Major [= Hp.Ma.]. (1903). In *Opera*. Vol. 3. Oxford: Clarendon Press.

Prior, A.N. (1967). Correspondence Theory of Truth. In Edwards, P. (ed.), *The Encyclopedia of Philosophy*. Vol. 2. London: Routledge, pp. 223–232.

Quine, W.V.O. (1950/1974). *Methods of Logic*. 3rd edition. London: Routledge.

Rumfitt, I. (2011) Truth and the Determination of Content: Variations on Themes from Frege's 'Logische Untersuchungen'. *Grazer Philosophische Studien* 82: 3–48.

Strawson, P.F. (1976). Knowledge and Truth. *Indian Philosophical Quarterly* 3(3): 273–282.

William of Ockham. (1978). Expositio in librum Peri hermeneias Aristotelis. In *Opera Philosophica* [= OPh II]. Vol. 2. St. Bonaventure, NY: The Franciscan Institute.

Williamson, T. (1999). Truth-Makers and the Converse Barcan Formula. *Dialectica* 53: 253–270.

Williamson, T. (2003). Everything. *Philosophical Perspectives* 17: 415–465.

Wilson, W.K. (1990). Some Reflections on the Prosentential Theory of Truth. In Dunn, M.J. and Gupta, A. (eds.), *Truth and Consequences*. Dordrecht: Kluwer Academic Publishers, pp. 19–32.

Wittgenstein, L. (1921/1961). *Tractatus Logico-Philosophicus* [= TLP]. London: Routledge.

Wittgenstein, L. (1931–32/1974). *Philosophical Grammar* [= PG]. Oxford: Blackwell.

Wittgenstein, L. (1953/2009). *Philosophical Investigations* [= PI]. 4th revised edition. Edited by P.M.S. Hacker and J. Schulte. Oxford: Wiley-Blackwell.

von Wright, G.H. (1984). *Truth, Knowledge and Modality*. Oxford: Blackwell.

von Wright, G.H. (1986). Truth, Negation and Contradiction. *Synthese* 66(1): 3–14.

von Wright, G.H. (1996). *Six Essays in Philosophical Logic*. Helsinki: Philosophical Society of Finland.

6 What can we learn meta-philosophically from how philosophers actually proceed in the philosophy of religion?

Ansgar Beckermann

1. Hanjo Glock is convinced that there is a real difference between philosophy and the empirical sciences. His 'approach distinguishes the conceptual issues of philosophy from the factual issues of science' (Glock 2017, 77). The area of philosophy which he concentrates on to illustrate his view concerns the question of animal minds, i.e., to be more precise, the two questions: 1. Do some non-human animals have mental properties? 2. If so, which animals have which mental properties? According to Glock, the answers to these questions depend on two factors:

> [O]n the one hand, they depend on contingent facts about animals to be established by empirical science, whether through observations in the field or experiments in the laboratory. On the other hand, they depend on what one makes of heavily contested concepts like that of a mind, of thought, rationality, belief, consciousness, desire, perception, sensation, intention, behaviour, action, emotion, and so on. (Glock 2017, 78)

Thus, there are two kinds of questions—conceptual and factual questions. Conceptual questions are questions like: What do the concepts *perception*, *memory*, *intention*, *behaviour*, and *action* mean? What are the conditions under which we can correctly say that a being perceives something, remembers something, decides to do something, feels something? Factual questions are questions like: Do animals feel pain? And if so, which animals do feel pain? Can a dog think about what to do next and then make a decision? Do squirrels remember where they buried their nuts? In my view, Glock holds that the mentioned factual questions can be answered only if first the corresponding conceptual questions have been answered. This is the job of philosophy. After that is done, the questions are handed over to the empirical sciences. For it is their task to find out all the contingent facts on the basis of which it can be decided which mental concepts apply to which animals. But as I will argue, sometimes even philosophers try to answer factual questions—with

DOI: 10.4324/9781003202929-9

and without the assistance of the empirical sciences, and at least in part without previously doing much conceptual analysis.

However, I do not want to argue with Glock with regard to his favourite philosophical question—the question of animal minds. Instead, I shall shift the focus towards another area of philosophy—an area in which I have recently been engaged myself, the philosophy of religion. My question is what we can learn in terms of meta-philosophy when we look more closely at which questions are dealt with by philosophers of religion and how they try to answer these questions. I am well aware that this will not allow us to draw conclusions for all areas of philosophy; in other areas—philosophy of mind, epistemology, or philosophy of language—philosophers surely have very different approaches. I do hope, however, that a closer look at philosophy of religion can teach us something about what philosophy is and how philosophers proceed.

In the following, I will start by giving an overview of the key points of debate. I will first focus on the arguments for the existence of God, then turn to the debates about the concept of God, the problem of suffering, and, finally, to questions that emerge from religious, and in particular mystical, experiences.

2. There are several versions of every argument for the existence of God. All *cosmological* arguments are based on relatively simple and undeniable facts—for example, the existence of the world or the existence of movement and causality. Let us take a look at two of these arguments. The first one proceeds from the well-known truth that there are contingent facts, facts that actually obtain but which could also have failed to obtain (Leibniz); the second argument assumes that our universe has a beginning (*Kalam* cosmological argument).

Leibniz's argument[1]
(A1) Every contingent fact has a sufficient reason as to why it obtains.
(A2) There is a contingent fact that includes all other contingent facts.
(A3) This fact must also have a sufficient reason.
(A4) Thus, there must be a sufficient reason that is not contingent but necessary, and this reason is what is called 'God'.

Kalam cosmological argument[2]
(B1) Whatever begins to exist has a cause.
(B2) The universe has begun to exist at some point in time.
(B3) Thus, the universe has a cause.
(B4) Since the universe is all of space-time reality, including all matter and energy, this cause, being distinct from the universe, must be a non-physical, immaterial being beyond space and time.
(B5) The only non-physical, immaterial beings beyond space and time are abstract objects like numbers and unembodied minds.

(B6) Abstract objects cannot cause anything.
(B7) Therefore, the cause of the universe must be a transcendent mind 'which is what believers understand God to be' (Craig 2010, 59).

Obviously, all of these arguments are based on premises that are not immediately self-evident to everyone. Therefore, their advocates have attempted to add further arguments to support, for example, the principle of sufficient reason (A1), the principle of causality (B1), or premise (B2).

What speaks in favour of thesis (A1) or thesis (B1)? The assumption that (B1) is self-evident can be found, for example, in Descartes:

> For it is very evident by the natural light [. . .] that nothing comes from nothing [. . .].
>
> (Descartes 1985, 199)

That (B1) is an assumption every reasonable person has to make was advocated, for example, by Copleston in his discussion with Russell (Russell and Copleston 1948). Other philosophers also have tried to *prove* the principle of causality in a more direct way.

In Leibniz, for instance, the following argument for (A1) can be found: Each thing (*res*) has necessary conditions for its existence; if only one of these conditions is not fulfilled, this thing does not exist. Thus, if something exists, then all its necessary conditions are fulfilled. And if all necessary conditions are fulfilled, the thing exists. If it does not exist, it does not exist because one of the necessary conditions is not fulfilled. All necessary conditions together are therefore a sufficient reason. Thus, there is a sufficient reason for everything that exists (Leibniz 1990, 483).

With regard to (B1), Hume attributes the following argument to Locke (Hume 1978, 81): If there is no cause for the beginning of A's existence, then A came into being from *nothing*, so *nothing* is then the cause of A; however, *nothing* cannot be a cause. Therefore, everything has something as its cause which is not nothing. Hume criticises this argument: If one starts from the assumption that everything has a cause, one inevitably comes to the conclusion (after having excluded all other causes) that A itself or *nothing* is the cause for the beginning of A's existence. But what is at issue is precisely the question whether everything in fact has a cause; this should not already be presupposed. One could add: If one understood the sentence 'Nothing is the cause for A to begin to exist' as saying 'The beginning of A's existence has a cause, namely, nothing', one would commit a severe logical mistake. This is so because the expression 'nothing' is not a name; it does not refer to an obscure object that is called 'nothing'. 'Nothing' is a quantifier and 'Nothing is the cause for A to begin to exist' does not state anything other than '*There is no cause* for the beginning of A's existence'.

Kant agrees with Hume that the principle of causality is not an analytic truth. It is, however, in his view, true and even *a priori* true. For, according to Kant, objective experience would be impossible if the

following statement were not true: 'All alterations occur in accordance with the law of the connection of cause and effect' (Kant 1998, A 189/B 232, 286). In particular, it would then be impossible to determine the objective, temporal sequence of events in the world of appearances. Since objective experience is possible, however, the statement must be true, according to Kant, at least in the phenomenal world, the world as it appears in our experience.

Even recently, philosophers have tried to prove the truth of the principle of causality. Craig and Sinclair, for example, write:

> First and foremost, the principle is rooted in the metaphysical intuition that something cannot come into being from nothing. For to come into existence without a cause of any sort is to come into being from nothing. To suggest that things could just pop into being uncaused out of nothing is to quit doing serious metaphysics and to resort to magic.
>
> (Craig and Sinclair 2012, 182)

The decisive point here is that Craig and Sinclair seem to take the metaphorical use of 'pop into being from nothing' literally, just as Locke seemed to do. This becomes apparent also in Craig's response to the argument that in particle physics, there are time and again particles that originate from nothing:

> This skeptical response represents a deliberate abuse of science. The theories in question have to do with particles originating as a fluctuation of the energy contained in the vacuum. The vacuum in modern physics is not what the layman understands by 'vacuum', namely, nothing. Rather in physics the vacuum is a sea of fluctuating energy governed by physical laws and having a physical structure.
>
> (Craig 2010, 76)

It seems to me, however, that the misunderstanding here is on the side of Craig and Sinclair. Who would really claim that the respective particles literally originate from nothing? The claim is simply that they come into being without a sufficient cause. *To come into being without a sufficient cause* is not the same as *to originate from nothing*. If Craig claims that nobody honestly believes that a horse, for instance, or an Inuit village comes into being without a cause, he may be right. But is the following scenario not at least conceivable? John enters a room in which he has never been before. Suddenly, he notices that a red ball is starting to emerge in a corner. At first, it is a very small ball that gradually becomes bigger until it has the size of a tennis ball. He then observes the same phenomenon in another corner and so on. And just as they emerged, the balls disappear again after a while. It is later explained to him that this is a so

far inexplicable phenomenon. And what is more, there is strong evidence that the balls, indeed, come into being without any sufficient cause. This does *not* mean, however, that they come into being from nothing.

On the other hand, proponents of the principle of causality often point out that this principle is empirically well proven. Craig and Sinclair write:

> Finally, [(B1)] is constantly confirmed in our experience. Scientific naturalists thus have the strongest of motivations to accept it.
>
> (Craig and Sinclair 2012, 187)

And what about the claim that (B1) is an assumption which every reasonable person and especially scientists have to make if they want to learn something about the world? As already mentioned, this claim was advocated by Copleston in 1948 in a BBC debate with Russell, but also by William Craig in October 2015 in Munich in a debate with me on the question *Does God Exist*? In my second response to Craig, I answered similarly to Russell:

> The assumption that there is an explanation for *everything* indeed does *not* belong to the standard assumptions of modern science. Only the methodological rule is valid: Try to find an explanation for the existence of every object as long as possible. But this rule can have the consequence that, at the end of long scientific investigations, we give up on this endeavour and conclude: In this case we cannot find an explanation; thus, this suggests that there is no explanation. I think precisely this is what happened in the development of quantum mechanics.
>
> (Beckermann and Craig 2015, second response, my translation)

Let us draw a first conclusion on the basis of what we have seen so far. The debate about whether one can conclude God's existence from the existence of the universe, the existence of causal chains or contingent facts, is not primarily about analysing concepts but about the factual question: Does God exist? Analyses of concepts do not play an important role in this debate. The core of the debate rather consists of arguments which aim to show that it is indeed possible to draw such conclusions. And the debate itself is then about how good these arguments are. Are the premises as plausible as the proponents of the arguments believe? And do the conclusions really follow from the premises?

The *ontological* arguments for the existence of God may be different since they start from theses for which it is not really clear whether they are factual or conceptual theses. Think, for example, of Anselm's argument:

> (C1) The *being than which no greater can be conceived* exists in the understanding.

(C2) The *being than which no greater can be conceived* exists only in the understanding and not in reality.
(C3) If something exists in the understanding, it is conceivable that it exists in reality as well.
(C4) It is greater to exist in reality than to exist merely in the understanding.
(C5) The *being than which no greater can be conceived* is something than which something greater can be conceived.

(C2) thus leads to a contradiction; therefore, (C2) is false and must be replaced by:

(C6) The *being than which no greater can be conceived* not only exists in the understanding but also in reality.

Even this argument, however, contains premises for which it is not entirely clear what they mean and why we should accept them. For example, what is meant by premise (C1)? Is existence a property of objects or, as Frege suggested, a property of concepts? And if F is a property, is an F that exists greater than an F that does not? I find it difficult to describe this debate with common categories. Is this an analysis of the concept of existence? Perhaps, but in any case, empirical arguments do not seem to play a role in this debate.

This is quite different in the so-called *teleological* arguments for the existence of God, which for a long time had the most supporters. The starting point for the traditional versions of these arguments is the undeniable fact that there are functional systems in the world (e.g., animals), i.e., beings that consist of functionally coordinated parts that give the beings extraordinary abilities, such as the ability to preserve themselves. (Whether the world as a whole has such a functional structure has always been subject to controversy.) How could such beings come into existence? Some functional systems (e.g., machines) are created by human engineers, i.e., intelligent beings. Can functional systems also come into existence in other ways, i.e., not by being designed and manufactured by some intelligent being? To many, this seemed unthinkable for a long time. It was an empirical theory, Darwin's theory of evolution, that changed everything. This theory made it clear that there are physical mechanisms (mechanisms in which intelligent engineers do not play a role) which ensure the gradual development of simple systems into more complex functional systems with increasingly extraordinary abilities. This insight, however, did not end the debate. Some argued that there are structures within living beings that could not have come into existence in an evolutionary way. In general, all this led to an interesting debate about which functional systems can be assumed to be the design of an intelligent engineer and which are the result of evolutionary mechanisms. In the favourite example of advocates of teleological arguments—the clock

that one finds in the forest—it is safe to assume that the clock was created by a watchmaker because, first, clocks do not have the ability to reproduce (a necessary element of evolutionary mechanisms). Second, the clock does not have an evolutionary history that could explain its existence. This seems to be completely different with regard to living beings.

A second version of teleological arguments is the so-called argument from fine-tuning. According to modern physics, there are a number of fundamental physical constants in our universe whose values could, as far as we know, have been different from what they really are. Theoretical calculations have shown that the universe would have developed quite differently if the values of some of the physical constants had been only slightly different. In these cases, there would not have been an expansion of the universe, no stable atoms, thus no life, no evolution etc. Why is it that the fundamental physical constants have exactly those values that enabled the existence of galaxies, the sun, planets, living beings, and, finally, mankind? Advocates of the argument from fine-tuning claim: The only plausible explanation is the assumption that an intelligent being created the world and designed these physical constants. The main argument for this thesis is: That the physical constants are precisely the way they are in order to enable the existence of galaxies, the sun, planets, living beings and mankind is so unlikely that it cannot be explained as a coincidence. Here, too, an interesting debate emerged. Of what unlikely events is it legitimate to assume that they are the result of an intervention by intelligent beings? And of what unlikely events is it appropriate to assume coincidence? If someone wins the lottery, it is usually legitimate to assume that it was purely coincidental that this person won. If, however, the same members of a Mafia family win the lottery over and over again, it is natural to assume manipulation. Where exactly is the difference between these two cases (see Weidemann 2019; Beckermann 2019b)?

Conceptual analysis plays a central role in philosophy of religion when the focus is on how the concepts *religion* and *God* (or *supernatural being*) are to be understood. These debates, if I am correct, are not primarily about analysing how these concepts are used but rather about what is the best way to understand these concepts. I sympathise with von Kutschera's characterisation of *religion*:

> A *religion* is the development of the awareness of a reality transcending the empirical world, the world of our normal experiences [. . .] In any case, in comparison to the empirical reality, the transcendent reality is viewed as the determining, larger, more powerful, and more significant reality along which we as humans should orient ourselves if our life is not to fail.
>
> (von Kutschera 2014, 141, my translation)

This characterisation seems to rather be an explication in the sense of Carnap than to be based on an analysis of how the term 'religion' is used in everyday lingo.

Regarding the concept of God, there are different suggestions as to how it should be understood. Norbert Hoerster follows the understanding of 'God' according to the Abrahamic religions:

> In a theistic understanding, God is the unique, eternal, personal, and non-physical, most perfect being who created, preserves, and directs the world.
>
> (Hoerster 2005, 13, my translation)

As already mentioned, in Anselm we find the formula: 'God is the *being than which no greater can be conceived*'. Other philosophers conceive of God as, for example, 'a supernatural being that humans should honour or the highest reality, the source of everything else, perfect and worth worshipping'.[3] All these suggestions are, however, obviously explicitly or implicitly determined by our Jewish-Christian culture. But, of course, the Greek gods are gods too. However, none of the properties described by Hoerster as characteristic for God seem to apply to them. They are neither unique nor highly perfect; they have not even created the world. They are personal beings, though, that differ from us humans in two points—they are much more powerful, and they are immortal. It is unclear whether they are non-physical; they are invisible most of the time but can also appear and act in many different shapes throughout the world. In my opinion, these different perspectives are not very interesting from a meta-philosophical point of view. To me, it does not seem appropriate to distinguish between 'right' and 'wrong' concepts of God.

The debate about the concept of God becomes more interesting when the focus is on whether the understanding of God as an omnipotent, omniscient, and omnibenevolent being is coherent at all. This is even more relevant, as some philosophers ascribe four additional characteristics to God—simplicity, eternity, immutability, and impassibility. Finally, the question is whether the concept of a necessary being is coherent at all. There are obviously tensions between the alleged characteristics of God and between these characteristics and other assumptions—for example, that God actively intervenes in the world at certain times or that some human actions are free in the libertarian sense. I myself have great doubts that non-physical persons can exist at all.

Two other central areas of philosophy of religion concern the problem of suffering and the role of religious experiences. It is obvious that there is a horrific amount of suffering in the world. Already in antiquity, the question was raised how this is possible if there is an omnipotent, omniscient, and omnibenevolent God who loves his creatures more than anything else. The problem of suffering has two forms—the logical problem, where the focus is on whether the characteristic features of God are logically inconsistent with the existence of suffering; and the evidential problem, which does not assume a logical inconsistency but

which claims that the existence of suffering makes it very improbable that a God with characteristics as proposed by the Abrahamic religions exists.

The classic formulation of the logical problem can be found in Lactantius: Either God wants to end the suffering and cannot, then he is powerless; or he can end the suffering but does not want to, then he is malevolent; or he does not want to end the suffering and cannot, then he is malevolent and powerless; or he wants to end the suffering and can, but where does the suffering then come from? (Lactantius 2018: *De iradei*, ch. 13) Lactantius' second point ('if God can end the suffering but does not want to, then he is malevolent'), however, is not as self-evident as it may seem at first sight. And the same is true for the related statement:

> (1) An omnipotent, omniscient, and omnibenevolent God prevents all suffering.

For it takes only a brief moment to realise that an omnipotent, omniscient, and omnibenevolent God would only prevent the suffering that is unnecessary for the accomplishment of more important goals. Whoever concludes that God does not exist because suffering exists would, therefore, not only have to claim that

> (2) There is a horrific amount of suffering in the world,

but also that

> (3) There is pointless suffering in the world, i.e., suffering that is not necessary for the accomplishment of more important goals.

And while (2) cannot be disputed, this does not seem to apply to (3) in the same way. It seems at least possible that (3) is wrong. However, if one asks what could be the more important goals for the sake of which God lets devastating catastrophes such as the earthquake of Lisbon or the tsunami in 2004 happen, then one is likely to hear the response: The Lord works in unfathomable ways. Is this a satisfying answer? And, even if it is possible that (3) is wrong: If a little fawn pitifully dies because its mother has been shot and it is caught in a trap, how probable is it that there is a more important goal for the sake of which God does not intervene?

Proponents of the logical version of the problem of suffering also have another way to challenge theism. Instead of claiming (3), they can hold that

> (4) There is suffering in the world where those that suffer are instrumentalised or violated in their rights.

In the story of Job, is it not the case that Job's children are inflicted with suffering in order to test Job's faith? Are the children then not instrumentalised? Or are they not at least violated in their rights? If that were the case, no goal could justify the suffering of these children. Even in the case of (4), one could, of course, argue that it is at least possible that it is wrong. But again: Does not rather a lot speak for the truth of (4)?

The evidential problem is often depicted in ways like this: Statements such as (3) are not undeniable, but nonetheless, it is very probable that they are true, and therefore it is also very probable that the conclusion is true, that is, that there is no God in the sense of the Abrahamic religions. In my opinion, however, it seems more appropriate to distinguish between the truth of the premises and the type of the inference. The latter is about whether the conclusion follows from the premises with deductive certainty or only with a certain probability; the former is about whether the premises are undisputable or merely probably true. I would therefore formulate the logical problem of suffering in the following way: There are undisputable or very likely facts that, with deductive certainty, lead to the conclusion that there is no God in the sense of the Abrahamic religions. Accordingly, for the evidential problem, the following would hold true: There are undisputable or very likely facts which make it very likely that there is no God in the sense of the Abrahamic religions.

One version of the evidential problem is the following. If we assume that supernatural beings have created our world, what does the design of this world tell us with great likelihood about the characteristics of these supernatural beings? If we look at the world around us, we undeniably see a lot of beauty and things worth admiring. On the other hand, this world is also truly imperfect. There are terrible diseases and natural disasters; people commit the most horrific crimes, and hundreds of thousands have to endure terrible pain and unfathomable suffering. Thus, the world has two faces—beautiful and impressive, on the one hand, and evil and terrible, on the other. Which assumption is more in harmony with this fact—the assumption that this world was created by an omnipotent, omniscient, and omnibenevolent God, or one of the assumptions that Hume proposes as alternatives?

> [Perhaps this world] was only the first rude essay of some infant deity who afterwards abandoned it, ashamed of his lame performance; it is the work only of some dependent, inferior deity, and is the object of derision to his superiors; it is the production of old age and dotage in some superannuated deity; and ever since his death has run on at adventures, from the first impulse and active force which it received from him.
>
> (Hume 1970, 53)

Or perhaps the assumption proposed by Manichaeism that there are two principles in the supernatural world—one good, one evil—that are in conflict with each other?

Let us finally turn to the problem of religious and, in particular, mystical experiences. All of us are constantly subject to sensory (visual, auditive, olfactory) experiences that inform us about our world. But some people make experiences in certain situations that go beyond our everyday environment and point to a different kind of reality. At the end of a meditation, for example, some may experience what is often described as 'enlightenment'. My brother-in-law, a Zen master, once told me what is characteristic of such an 'enlightenment':

> The breakthrough can occur in an instance when you are completely drained from thoughts and feelings [. . .]. You see a world in which everything has disappeared, in which there is no distinction between subject and object. You are inside and outside at the same time. There is only one thing, my actual being. This view lasts for seconds, at the most a few minutes. There is no fear of the coming of 'emptiness' [. . .] The entire experience is astonishment about this world in which obviously all common thoughts and feelings have disappeared. Everything has melted into this one thing [. . .] At the same time, this world as we see it every day is absurdly considered to be good [. . .] In the case of deep enlightenment, there is always a fear of the great, of the unknown combined with the unspecific feeling that everything that you have learned or experienced is wrong. In view of this greatness you become small and humble, no matter how great the effort was, pride does not occur.
>
> (Personal conversation)

The reality with which one is apparently confronted in such experiences is, thus, very different than the world of our everyday experience. One cannot really say what it is like, but it is obviously overwhelming. The decisive philosophical question, however, is: Do mystical experiences show us a reality the way it is? Or are they hallucinatory?—Do they pretend a reality that, in fact, does not exist?

William Alston made an important attempt at arguing for the reliability of mystical experiences.[4] According to Alston, an individual belief is justified if it is based on a reliable doxastic practice. Doxastic practices include the practice 'to form perceptual beliefs on the basis of perceptual experiences, beliefs about one's own consciousness on the basis of introspection, inferential beliefs on the basis of deductive and inductive inferences, and memory beliefs on the basis of memories' (Jäger 2009, 402, my translation). But they also include the practice to form beliefs about a transcendent reality on the basis of mystical experiences. The decisive question, thus, is: Which of these practices are reliable?

One of Alston's central claims is that the epistemic practices of the standard package can be proven to be reliable only in a circular manner. For the assessment of the mystical practice, this has the following consequence. If it, too, can only be justified in a circular manner, it would, in this respect, not be inferior to the other practices of the standard package. Even if doxastic practices cannot be proven reliable in a non-circular manner, it is according to Alston nonetheless rational to trust them. With regard to the perceptual practice, he argues that it is definitely rational *in a practical sense* to follow at least socially well-established practices. This is so because, first, it is impossible to give up beliefs that were developed this way. Second, even if we had the choice between the perceptual and alternative practices, we would never have a good reason to prefer one of the alternative practices to the perceptual practice.

The upshot of these considerations is that, in Alston's view, there are no significant differences between the perceptual and mystical practice that would justify considering one reliable and the other one not. While Alston does respond to some of the arguments that have been brought forth against this thesis, he, in my opinion, does not pay enough attention to the central counterargument.

If we look at the reliability of sensory experiences, we know due to our everyday experience and due to scientific analyses that there are normal circumstances in which these experiences inform us reliably about our environment. And we know that there are also extraordinary circumstances in which this is not the case. It cannot be denied, however, that mystical experiences only occur under very special conditions. Meditation, strict fasting, or the intake of drugs, which sometimes also leads to such experiences, certainly do not belong to normal circumstances of our system of perception. And this is a strong reason for assuming that mystical experiences are rather hallucinations than veridical perceptions. Bertrand Russell succinctly summarised this argument:

> From a scientific point of view, we can make no distinction between the man who eats little and sees heaven and the man who drinks much and sees snakes. Each is in an abnormal physical condition, and therefore has abnormal perceptions.
>
> (Russell 1997, 188)

Moreover, we are living beings with a biological make-up. And we now know a lot about the biological mechanisms that underlie our perceptual abilities. And as a result of this knowledge, there are even more doubts about the reliability of mystical experiences. (See, e.g., Beckermann 2013, sect. 8.2.)

3. So far, I have tried to give an overview of some, by no means all, main topics of the philosophy of religion. What can we learn from

this meta-philosophically? First of all, this branch of philosophy, as it is conducted (again) today, is not about therapy: Today, hardly anyone attempts to reveal the problems of philosophy of religion as pseudo-problems through linguistic analysis anymore. These questions and problems have turned out to be real questions and problems that cannot be discussed away. They are about factual issues, and in particular about the questions: Does a supernatural world exist? And if so, what does it look like?

Second, one has to admit that the analysis of concepts does only play a minor role in the philosophy of religion. Of course, there is a debate about whether omnipotence, omniscience, omnibenevolence, simplicity, eternity, immutability, and impassibility form a coherent set of properties and whether the concept of a necessary being makes any sense at all. This debate, however, has a different significance than the debate on what the mental concepts of thought, rationality, belief, consciousness, desire, perception, sensation, and intention amount to with regard to the question of animal minds. Philosophy of religion does not start by asking what the decisive features of supernatural beings are, and then pass the question on to the empirical sciences whether there are beings that have these features. It rather takes the concepts of a transcendent world and of supernatural beings as given, and directly attempts to try to find out whether there is such a world and whether there are such beings. But why is this so? Why is it not the task of the empirical sciences to answer these questions?

At least some authors have claimed that the sciences as a matter of principle cannot say anything about the world of the supernatural. Kurt Bayertz, for example, argues:

> In the framework of the natural sciences, [the principle of immanence, according to which all real phenomena can be explained by reference to material objects and processes,] functions as a methodological rule, that only those things can be taken into account which can be 'explained naturally'. Phenomena that elude exact empirical analysis are simply not in the scope of scientific research. Whether they exist in reality or not is a question that cannot even be asked in the framework of the natural sciences, let alone be answered. The experimental method cannot help in detecting anything about spirits and gods; but this does not mean, of course, that spirits and gods do not exist. A materialistic interpretation of the essence of the world thus necessarily goes beyond what can legitimately be claimed in the framework of the natural sciences.
>
> (Bayertz 2007, 57–58, my translation)

I consider this view to be entirely wrong. Many who believe in a transcendent world and in supernatural beings also believe that these beings are able to influence their lives in both positive and negative ways. Before

undertaking a journey or starting a business or in case of a serious illness, they, therefore, pray to the beings they believe in or offer sacrifices or take vows. And it is, of course, an empirical question whether these prayers and sacrifices are successful. Many supernatural beings are regarded to be especially touchy in matters of honour; they are believed to tend to immediately punish people who do not pay sufficient respect. And again, it is an empirical question whether people who act insubordinately suffer more than others. It is also an empirical question whether some people regularly conduct actions that cannot be explained in a natural way (rainmakers, witch doctors).[5] In general, one can ask whether there are phenomena in our world which can only or best be explained by the intervention of supernatural beings. Thus, why is it that many believe that we cannot pass on to the empirical sciences the questions whether there is such a world and whether there are such beings?

I have the impression that nearly everyone in the debate now agrees that there are *de facto* no phenomena in our world which can only or best be explained by the intervention of supernatural beings. There are no miracle healings, no resurrections of the dead, no miraculous rescues. In general, there are no 'normal' empirical phenomena that speak in favour of the existence of a transcendent world or the existence of supernatural beings. This, however, does not terminate the debate. For many think that there may be other ways to argue for or against the existence of such a world and of such beings.

One such way consists in taking a step back, reconsidering some rather general facts, and asking whether these facts might speak for or against the existence of a transcendent world. This strategy forms the background of most arguments for the existence of God, with the exception of the ontological arguments. And it also forms the background of the argument from suffering.

Ontological arguments are special. One of these arguments claims that the statement 'God exists' is analytically true because the assumption of the opposite contains a contradiction. This thesis has led to interesting discussions about the nature of existential propositions and, in particular, about whether there can be analytically true existential propositions at all. I find it difficult to categorise this debate metaphilosophically. Is this an attempt to ascertain essential features of the world by philosophical means? Or is it a matter of conceptual analysis? Or a matter of a scientific investigation?[6] It does not seem to be the latter, because how would one want to find out whether there are existential propositions that are analytically true by using normal scientific methods? This question can perhaps be answered by searching for examples. Are there any provable existential propositions in mathematics, such as 'There is a smallest natural number'? If yes, what characterises them? Does 'God exists' belong to the group of existential propositions that can be proven?

Things are different with regard to the other classic arguments for the existence of God as well as to the problem of suffering. Here, the starting points are (allegedly) undeniable facts—the fact that our universe has begun to exist at some point in time; the fact that there are functional beings (animals) in our world which were not created by humans; the fact that the physical constants are such that they enable the formation of galaxies, the sun, planets, living beings, and humans; and the fact that there is a horrific amount of (pointless) suffering in the world. The basic questions then always are: Are these really facts? And if so, what can we conclude from these facts?

The starting point of the *Kalam* cosmological argument is premise (B2), 'The universe has begun to exist at some point in time'. In their long article 'The *Kalam* Cosmological Argument', Craig and Sinclair use an interesting mix of arguments to argue for the truth of this premise. On the one hand, they base their considerations on primarily philosophical arguments for the thesis that there can be no actual infinite. On the other hand, a large part of their account is dedicated to an extensive discussion of the main contemporary scientific cosmologies. They conclude that these cosmologies suggest that our universe, indeed, has begun to exist at some point in time.

The most controversial point of the *Kalam* argument, however, is premise (B1)—the principle of causality. As sketched above, there are four main strategies to argue for this premise: First, the claim that the principle is self-evident; second, the claim that the principle is an assumption that every reasonable person and especially every scientist has to make if she/he wants to learn something about the world; third, the attempt to 'prove' this principle in one way or another; and finally, the claim that the principle is empirically well confirmed. Thus, there are very different ways in which philosophers tried to argue for the truth of (B1), some of which rely on empirical findings, while others rest on quite different considerations. The question whether (B1) is self-evident is particularly interesting, since here views differ tremendously. Some find this so obvious that they tend to declare philosophers who doubt it crazy. Others like myself do not have a problem imagining situations in which things begin to exist without any sufficient cause. This seems to be a case of unsolvable philosophical dissent.

The classic teleological arguments for the existence of God are also interesting meta-philosophically. One version of these arguments is essentially about how complex functional systems can come into existence. Only by being planned and created by intelligent beings? Or can they also come into existence through purely physical mechanisms in which intelligent engineers do not play a role? The traditional claim was:

> (5) Complex functional systems can only come into existence if they are designed and manufactured by intelligent beings.

Is this a philosophical or a scientific claim? In any case, it is a claim that has been debunked by the development of the sciences. In this context, the details of evolutionary theory are not decisive. What is decisive is that this theory has demonstrated what a natural mechanism driven by selection and mutation can look like that ensures the development from simple to more complex functional systems with increasingly extraordinary abilities. This robbed (5) of the intuitive plausibility it had for centuries or even millennia. As already mentioned, this was not the end of the debate as to whether the existence of living beings shows that there is an intelligent designer who planned and created them. There are now two potential explanations: Plants, non-human animals, and humans were designed and created by an intelligent being; or plants, non-human animals, and humans came into existence through a natural evolutionary process. Thus, it is still debatable which of these claims is correct. This dispute has, however, developed more and more into a scientific debate.

The fine-tuning debate has a different character. Two claims form the basis here:

> (6) The fundamental natural constants have values that have enabled the existence of galaxies, the sun, planets, living beings, and finally us humans.

And

> (7) That the fundamental natural constants have the values they have is highly unlikely.

(6) is obviously undeniable, while (7) can surely be disputed. Does it really make sense to speak of probabilities here? Aside from that, the core of the debate is, in this case, about two questions: 1. Of which unlikely events is it legitimate to assume that they result from the intervention of intelligent beings, and of which unlikely events is it sufficient to assume a coincidence? 2. Does the fact that the fundamental natural constants have the values they have belong to the former or the latter type? How can these questions be answered?

As we have seen, we are obviously able to make an intuitive distinction in some paradigmatic cases. We should, therefore, analyse what the cases have in common that we consider intuitively to belong to the first type, and what differentiates them from the cases that we intuitively consider to belong to the second type. Weidemann has made a suggestion based on considerations by Paul Horwich (Weidemann 2019). According to this proposal, an unlikely event belongs to the first type when it is not only unlikely but also surprising. An event E is surprising when: a) E is, given our basic assumptions C, very unlikely; b) there are alternative basic assumptions K, according to which E is much more likely; and c) the alternative basic assumptions K are (independently of E) implausible but not extremely

unlikely. I have commented on this proposal critically (Beckermann 2019b). But that is not important here. Rather, the question is: What kind of discussion is this? How can it be categorised meta-philosophically? I must confess that I know of no straight-forward answer to this question.

The problem of suffering has a by now well-known structure. It starts from an (in this case really) undeniable fact—that there is a horrific amount of suffering in the world—and asks what can be concluded from this fact. In the logical version of the problem of suffering, the question is what can be deductively concluded from this fact. It has become clear that it cannot be concluded with deductive certainty that there is no omnipotent, omniscient, and omnibenevolent God. Therefore, some have asked whether not only (2) 'There is a horrific amount of suffering in the world' but also (3) 'There is pointless suffering in the world' or (4) 'There is suffering in the world where those that suffer are instrumentalised or violated in their rights' are true. Of course, (3) and (4) may be wrong. But are they not at least very probable? And what does that mean for the power of the following argument?

> (4) There is suffering in the world where those that suffer are instrumentalised or violated in their rights.

Therefore:

> (8) There is no omnipotent, omniscient, and omnibenevolent god.

The same questions must be asked with regard to the argument:

> (9) There is a lot of beauty and a lot of things worth admiring in the world while, on the other hand, there are terrible diseases and natural disasters, people commit the most horrific crimes, and hundreds of thousands have to endure terrible pain and unfathomable suffering.

Therefore:

> (10) This world has not been created by an omnipotent, omniscient, and omnibenevolent god.

Again, we are confronted with the same questions. How well confirmed are the premises? Do the conclusions really follow—deductively or at least with a certain probability—from the premises?

The debate about the reliability of mystical experiences obviously reaches far beyond philosophy of religion. This problem can apparently not be dealt with independently of the question of what can be said about the reliability of our doxastic practices at all. Thus, we immediately enter the field of epistemology. As long as the focus is specifically

on the reliability of mystical experiences, the structure of the debate is relatively clear. At its core, it is about whether, as Alston claims, the mystical practice does not differ significantly from the perceptual practice. As far as I can tell, there is an interesting mix of philosophical and empirical arguments in this debate as well.

If we attempt to draw a brief conclusion, the following should have become clear: In terms of meta-philosophy, the debates in the philosophy of religion are very different from the debates concerning the question of animal minds as characterised by Glock. Glock's picture of a division of labour—science has the task to establish all the relevant contingent facts about animals, while the task of philosophy is to provide clarification with regard to the pertinent mental concepts—does not fit the debates in the philosophy of religion. The latter directly address the questions whether a supernatural world exists and if so, what it looks like. And philosophers have tried to answer these questions on their own. For there is a great consensus that empirical sciences are of little help here since there are obviously *de facto* no phenomena in our world that can only or best be explained by the intervention of supernatural beings.

Aside from trying to show that the statement 'God exists' is an analytic truth and from seeking refuge in mystical perception, there is a preferred strategy for adherents as well as for opponents of theism: Start from (allegedly) undeniable general facts and then try to show that these facts speak in favour of or against the existence of God. Thus, large parts of the discussion circle around what can be concluded from what. Can one conclude from the fact that our universe has come into existence that the universe has a cause, and that this cause is an intelligent, non-physical, and necessary being? Can one conclude from the fact that there are functional systems in the world not created by humans that there is a supernatural being who has created these functional systems? Can one conclude from the fact that the natural constants are the way they are that there is a supernatural being which has made sure the natural constants are the way they are? Can one conclude from the fact that in this world there is a horrific amount of suffering that this world was not created by an omnipotent, omniscient, and omnibenevolent God? To reach such conclusions, one would usually need additional assumptions, and these are also at the core of the discussion. To support the additional assumptions as well as the original premises, the motto seems to be: Take whatever you can get.

There is an obvious disadvantage of the sketched main strategy that is well captured by the slogan 'One person's modus ponens is the other's modus tollens'. There is foreseeable dissent. We have already seen that there is dissent with regard to the question of whether the principle of causality is self-evident—a dissent of which I cannot see how it could be resolved. And the same is true for all arguments I have discussed. One may think that the premises are true and that we therefore should accept the conclusions, or one may think that the conclusions are so implausible

that there must be something wrong with the premises. Furthermore, there may be dissent about the question whether the conclusion really follows from the premises. Thus, in this case philosophy cannot give final answers to the questions it addresses. It can only create clarity. It can make clear what assumptions you have to accept if you believe thesis A to be true and what you are committed to if you tend to believe non-A. I believe that this is a characteristic feature of most philosophy. In any case, it is what David Lewis says:

> Whether or not it would be nice to knock disagreeing philosophers down by sheer force of argument, it cannot be done. Philosophical theories are never refuted conclusively. (Or hardly ever. Gödel and Gettier may have done it.) The theory survives its refutation—at a price [. . .] [W]hen all is said and done, and all the tricky arguments and distinctions and counterexamples have been discovered, presumably we will still face the question which prices are worth paying, which theories are on balance credible, which are the unacceptably counterintuitive consequences and which are the acceptably counterintuitive ones. On this question we may still differ. And if all is indeed said and done, there will be no hope of discovering still further arguments to settle our differences.
>
> (Lewis 1983, x)

Notes

1 Leibniz 1989, cf. Pruss 2012, 25–26.
2 E.g. Craig and Sinclair 2012, 102; Craig 2010, ch. 4.
3 URL: https://de.wikipedia.org/wiki/Gott. Last accessed: 29 March 2022, my translation.
4 E.g. in Alston 1991. In the following, I orient myself by Christoph Jäger's considerations in his 2002 and 2009. See also Steup 1997.
5 I go into more detail on this in Beckermann 2019a.
6 On this, cf. Kutschera 2017.

References

Alston, W. (1991). *Perceiving God: The Epistemology of Religious Experience*. Ithaca, NY/London: Cornell University Press.

Bayertz, K. (2007). Was ist moderner Materialismus? In Bayertz, K., Gerhard, M. and Jaeschke, W. (eds.), *Weltanschauung, Philosophie und Naturwissenschaft im 19. Jahrhundert. Band 1: Der Materialismus-Streit*. Hamburg: Felix Meiner, pp. 50–70.

Beckermann, A. (2013). *Glaube*. Berlin/Boston: de Gruyter.

Beckermann, A. (2019a). Ein nüchterner Blick auf die Welt. In Jaster, R. and Schulte, P. (eds.), *Glaube und Rationalität*. Münster: mentis, pp. 15–30.

Beckermann, A. (2019b). Antworten. In Jaster, R. and Schulte, P. (eds.), *Glaube und Rationalität*. Münster: mentis, pp. 191–203.

Beckermann, A. and Craig, W. (2015). Gibt es Gott? https://www.reasonablefaith.org/videos/debates/craig-vs-beckermann-munich-germany/. Last accessed: 30 March 2022.

Craig, W. (2010). *On Guard: Defending Your Faith with Reason and Precision*. Colorado Springs, CO: David C. Cook.

Craig, W. and Sinclair, J.D. (2012). The *Kalam* Cosmological Argument. In Craig, W.L. and Moreland, J.P. (eds.), *The Blackwell Companion to Natural Theology*. Chichester: Wiley-Blackwell, pp. 101–201.

Descartes, R. (1985). *The Philosophical Writings of Descartes*. Vol. 1. Translated by J. Cottingham, R. Stoothoff, and D. Murdoch. Cambridge: Cambridge University Press.

Glock, H.-J. (2017). Impure Conceptual Analysis. In Overgaard, S. and d'Oro, G. (eds.), *The Cambridge Companion to Philosophical Methodology*. Cambridge: Cambridge University Press, pp. 83–107.

Hoerster, N. (2005). *Die Frage nach Gott*. München: C.H. Beck.

Hume, D. (1970). *Dialogues Concerning Natural Religion*. Edited by N. Pike. Indianapolis: Bobbs Merill.

Hume, D. (1978). *A Treatise of Human Nature*. Edited by L.A. Selb-Bigge, 2nd edition by P.H. Nidditch. Oxford: Clarendon Press.

Jäger, C. (2002). Religious Experience and Epistemic Justification: Alston on the Reliability of 'Mystical Perception'. In Moulines, C. and Niebergall, K.G. (eds.), *Argument und Analyse*. Paderborn: mentis, pp. 403–423.

Jäger, C. (2009). Epistemische Rationalität und Alstons Theorie 'mystischer Wahrnehmung'. *Zeitschrift für katholische Theologie* 131: 396–413.

Kant, I. (1998). *Critique of Pure Reason*. Edited and translated by Paul Guyer and A.W. Wood. Cambridge: Cambridge University Press.

Lactantius, C.F. (2018) *Opera Omnia*. Vol. 2. London: Forgotten Books.

Leibniz, G.W. (1989). On the Ultimate Origination of Things. In Leibniz, G.W. *Philosophical Essays*. Indianapolis, IN/Cambridge: Hackett Publishing Company, pp. 149–154.

Leibniz, G.W. (1990). *Philosophische Schriften. Sechste Reihe: Band 2*. Edited by Leibniz-Forschungsstelle Münster. Berlin: Akademie-Verlag.

Lewis, D. (1983). *Philosophical Papers*. Vol. 1. New York, NY/Oxford: Oxford University Press.

Pruss, A.R. (2012). The Leibnizian Cosmological Argument. In Craig, W.L. and Moreland, J.P. (eds.), *The Blackwell Companion to Natural Theology*. Chichester: Wiley-Blackwell, pp. 24–100.

Russell, B. (1997). *Religion and Science*. Oxford/New York, NY: Oxford University Press.

Russell, B. and Copleston, F.C. (1948). *A Debate on the Existence of God*. http://www.scandalon.co.uk/philosophy/cosmological_radio.htm. Last accessed: 30 March 2022.

Steup, M. (1997). William Alston, Perceiving God. The Epistemology of Religious Experience. *Nous* 31: 408–420.

von Kutschera, F. (2014). *Drei Formen des Bewusstseins*. Münster: mentis.

Weidemann, C. (2019). Dinge, die unwahrscheinlich sind, passieren einfach. In: Jaster, R. and Schulte, P. (eds.), *Glaube und Rationalität*. Münster: mentis, pp. 163–189.

7 More good news for the philosophical armchair

Christian Nimtz

Introduction: A division of labour?

Hanjo Glock holds that conceptual rather than factual issues 'constitute the proper domain of philosophy' (Glock 2017, 79; see also Glock 2010; Glock 2003, chs. 2–3). Yet he grants that tracing conceptual connections will often require substantial empirical input. Embracing a methodology he advertises as 'impure conceptual analysis', Glock (2017) programmatically insists that conceptual reflection on issues such as, say, the analysis of animal mind (Glock 2013, 2019) needs to be well informed about the pertinent empirical scientific problems, claims, and methods. The general idea is that when armchair philosophers do not engage with regular conceptual truths or with metaphysical necessities, but rather take up what I will dub *de facto scientific hypotheses*—i.e., empirical hypotheses that support counterfactuals and that fall within the purview of an empirical science—we find a characteristic 'division of labour' (Glock 2017, 81). The empirical sciences provide all the empirical evidence we need beyond the commonplaces we are familiar with anyway. The contribution of armchair philosophers, on the other hand, is purely theoretical: They exclusively contribute their distinctive prowess at 'elucidating existing concepts' (Glock 2017, 81).

Glock's methodological picture puts him in general agreement with recent defences of armchair philosophy from the opposing, naturalist end of the spectrum. Papineau (2014) and Nolan (2015), too, think that armchair philosophers can responsibly assess *de facto* scientific claims because philosophers are good at theorising. As Nolan stresses, the armchair is a good position to 'assess rival philosophical views in the light of general standards of theoretical virtue, such as relative simplicity, unificatory power, explanatory potential, and so on' (Nolan 2015, 215). Papineau, Nolan (2015, 223–224), and Glock (2017, 82) moreover agree that solving philosophical problems requires better armchair theorising, rather than more empirical data. As Papineau (2014, 167) puts it: 'The characteristic philosophical predicament is that we have all the data we could want, but still cannot see how to resolve our theoretical problems'.

DOI: 10.4324/9781003202929-10

I agree that the philosophical armchair is a good place to theorise, whether this amounts to tracing conceptual connections, as Glock thinks, or to assessing theoretical virtues, as Nolan and Papineau believe (see also Bach 2019). But I sense a problem with the meta-philosophical picture that Glock, Nolan, and Papineau so readily agree upon: Our philosophical practice goes beyond it. Within the philosophy of mind, on which I will focus, the philosophical armchair is not just treated as a place where empirical hypotheses are theoretically appraised. It is also treated as a place where evidence for such hypotheses is procured in the first place. Philosophers of mind habitually[1] accept *a posteriori* claims about the mind exclusively on the strength of a few imaginary armchair cases. Think, for example, of Block's (1995, 233, 246, fn.16) tale of super-blindsight, which he employs to warrant his views on access-consciousness, Armstrong's (1980, 59) case of the long-distance truck driver, which he uses to justify claims about how perceptual and introspective consciousness relate, or Dretske's (1988, 52–53) use of a nickel (heads uppermost) to represent Kareem Abdul-Jabbar, which Dretske appeals to in support of his theory of representation.[2]

I hold that armchair philosophy has more to contribute than theoretical prowess when it comes to assessing *de facto* scientific hypotheses about the mind. I will argue that armchair philosophers can legitimately procure evidence that is arguably apt to support such hypotheses. My argument proceeds via an in-depth case study. In his 'The Essential Indexical', John Perry commits to *a posteriori* hypotheses about human locating beliefs exclusively on the evidence gathered from a few imaginary scenarios. I argue that Perry is methodologically entitled to do so for two key reasons. First, Perry uses his scenarios in a distinctive way: He employs them as puzzle cases to identify *explananda* for an *a posteriori* theory of locating belief. Since a few well-chosen *explananda* can make all the difference in an abductive argument, as I argue, Perry's few cases can provide sufficient support for his empirical claims. Second, although the *explananda* Perry derives from his imaginary puzzles are, for all we know, not actually true, they could easily be true, and we know this. I argue that since known near-actual truths of this sort provide empirical backing, the data Perry derives from his imaginary cases are indeed suited to abductively support *a posteriori* hypotheses about human cognition.

As it turns out, then, the picture of a neat division of labour between philosophy and the empirical sciences so readily agreed upon by Glock, Nolan, and Papineau needs qualifying. So does Glock's key axiom that philosophy is essentially concerned with conceptual matters, if that claim is to be more than a triviality perfectly amenable to any philosophical naturalist averse to analytic truths.

Our case study introduced: Perry on shoppers, professors, hikers, and locating belief

John Perry's paper 'The Problem of the Essential Indexical' begins with his famed tales of the absent-minded shopper, the tardy professor, and the lost hiker. I take it as given that Perry's tales depict events he has imagined, rather than events he knows to have occurred:

> I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter, the trail became thicker. But I seemed unable to catch up. Finally, it dawned on me. I was the shopper I was trying to catch. I believed at the outset that the shopper with a torn sack was making a mess. And I was right. But I didn't believe that I was making a mess. That seems to be something I came to believe. And when I came to believe that, I stopped following the trail around the counter, and rearranged the torn sack in my cart. My change in beliefs seems to explain my change in behavior. My aim in this paper is to make a key point about the characterization of this change, and of beliefs in general.
>
> (Perry 2000, 27)

> [A] professor, who desires to attend the department meeting on time and believes correctly that it begins at noon, sits motionless in his office at that time. Suddenly, he begins to move. What explains his action? A change in belief. He believed all along that the department meeting starts at noon; he came to believe, as he would have put it, that it starts *now*.
>
> (Perry 2000, 28)

> The author of the book *Hiker's Guide to the Desolation Wilderness* stands in the wilderness beside Gilmore Lake, looking at the Mt. Tallac trail as it leaves the lake and climbs the mountain. He desires to leave the wilderness. He believes that the best way out from Gilmore Lake is to follow the Mt. Tallac trail up the mountain to Cathedral Peaks trail, on to the Floating Island trail, emerging at Spring Creek Tract Road. But he does not move. He is lost. He is not sure whether he is standing beside Gilmore Lake, looking at Mt. Tallac, or beside Clyde Lake looking at Jacks Peak, or beside Eagle Lake looking at one of the Maggie peaks. Then he begins to move along the Mt. Tallac trail. If asked, he would have explained the crucial change in his beliefs this way: 'I came to believe that this is the Mt. Tallac trail and that is Gilmore Lake'.
>
> (Perry 2000, 28)

Each of Perry's cases pivots on a *locating belief*, i.e. a belief 'where one is, when it is, and who one is' (Perry 2000, 29). In each example, a change in a locating belief, as witnessed by the subject's becoming willing to assert an indexical statement, comes with a distinctive change in intentional behaviour. On thinking '*I* am the shopper who is making a mess', the shopper stops and rearranges the sack in his cart. On thinking 'The meeting starts *now*', the professor gets up, and on thinking '*This* is the Mt. Tallac trail and *that* is Gilmore lake', the hiker starts moving.

Perry argues from these cases to a revisionist two-factor account of locating belief (Perry 2000, 38–41; see also Perry 2000, xi, chs. 19, 21; Perry 2019, chs. 2, 11). Philosophers standardly assume that a thinker *S* has a belief just in case *S* stands in the *believing* relation to a content *p* (see e.g. Fodor 1987, 17). Arguing that our theory needs to distinguish 'between *what* is believed and *how* it is believed' (Perry 2019, 7), Perry holds that locating beliefs comprise a second element besides their contents, which he dubs a 'belief-state' (Perry 2000, 39).

According to Perry, the content of a locating belief is a singular proposition containing the place, time, or subject the belief is about. The belief-state of a locating belief captures how the subject thinks of that place, time, or subject; it consists of the role that the thinker takes this object to play relative to herself. A belief-state may, for example, portray a place as playing the *here*-role, or a time as playing the *soon*-role. Perry's two-factor account allows that subjects who believe the same content may differ in their belief-states, and hence in their actions: 'When you and I both apprehend the thought that I am about to be attacked by a bear, we behave differently. I roll up in a ball, you run to get help' (Perry 2000, 19). Here, the two subjects believe of the same subject *S* that *S* is to be attacked by a bear. But the one subject thinks of *S* as playing the *I*-role, whereas the other subject thinks of *S* as playing the *you*-role. This explains their different actions.

Perry's two-factor account comes with a corollary about the role locating beliefs play in human intentional action. Let me call a mental state *truly indexical* if it comprises a non-eliminable indexical, perspectival, or *de se* element, be it an indexical content or an indexical element besides its content. And let me call the kind of action that the shopper, professor, and hiker perform a 'Perry-type action'. (Very roughly, an action is of this kind if it marks a behavioural change that occurs because of a change in a locating belief held by the subject, as witnessed by the subject's becoming willing to assert an indexical statement such as 'I am [. . .] ', or 'This is [. . .]') Then, we can capture the corollary Perry commits to by the following *Indexicality Lemma*: *Any Perry-type intentional action A performed by a normal human agent S is such that S's motives for A essentially comprise a truly indexical locating belief.*

The Indexicality Lemma might feel unduly modest. Perry is often cited as witness for the ambitious modal claim that *necessarily*, whenever *any*

intentional agent *S* performs *any* intentional action *A*, *S*'s motives for *A* essentially comprise a truly indexical belief (see Cappelen and Dever 2013, 37–39). I see no basis for this in Perry's texts. Neither does Perry himself. As his book-length 'Revisiting the Essential Indexical' (Perry 2019) makes abundantly clear, Perry means to capture 'contingently pervasive' (ibid., 111) truths about humans and other animals under 'anything approaching normal conditions' (ibid., 108). As he emphasises: 'My respect for the ability of philosophers to come up with possible counterexamples makes me very leery of claims of necessity' (ibid., 110). This is all for the better. As Cappelen and Dever (2013, 49–52) point out, the necessity reading just given is easily proven false by dreaming up possible intentional agents who lack locating beliefs altogether. For example, envisage intelligent beings who only perform mental acts of counting, or think of Lewis's (1979a) two gods who possess only third-personal information about their world. I see no reason to deny that such creatures are conceptually and metaphysically possible, or that they qualify as intentional agents. But if that is so, then intentional action is not by necessity tied to indexical belief.

Perry's empirical commitments and two methodological challenges

We have seen that Perry's proposal evades an easy refutation by modal counterexample—given, that is, that he argues for the Indexicality Lemma rather than for some necessity claim. However, we now have to concede that Perry aims to establish a *de facto* scientific hypothesis. As I explain presently, this creates methodological challenges to Perry's argument no less serious than an easy modal refutation.

Here is why the Indexicality Lemma qualifies as a *de facto* scientific hypothesis. First, the lemma is an *a posteriori* hypothesis. It assures us that human action of a certain type requires representational states with an in-built indexical component. It does so without committing to any specific implementation of (indexical) mental representations in humans (see von Eckardt 2012, sc. 2.3). Second, the Indexicality Lemma is metaphysically contingent. It holds true for human intentional agents like you and me under conditions like those we find in our actual world. Still, it supports the counterfactual that *if* a human agent *S* were to perform a Perry-type action *A*, then *S*'s motives for *A* *would* essentially comprise a truly indexical locating belief. So the lemma does have a modal consequence. But as it is typical for scientific claims, this consequence merely concerns worlds close to the actual world. Third, the Indexicality Lemma falls squarely into the purview of the empirical sciences of the mind. It claims that any correct theory of human mental representation needs to assign locating beliefs a non-eliminable indexical, perspectival, or *de se* element. Whether such an empirical (meta-)hypothesis proves

acceptable will depend on whether it is eventually borne out by empirical research, or so everyone agrees. Cognitive psychology, not philosophy, is the final arbiter of its truth.

Perry accepts a revisionist *a posteriori* two-factor account of human locating belief that commits him to a *de facto* scientific hypothesis in the form of his Indexicality Lemma, or so we have found. There is nothing puzzling in that. What is perplexing, however, is that Perry does not even consider data from the empirical sciences of the mind. Sticking to what he dubs 'amateur anthropology' (Perry 2019, 127) and 'amateur animal psychology' (ibid., 131), Perry exclusively relies on a few imaginary scenarios. This renders Perry's argument subject to two challenges.

The narrow range of evidence that Perry considers invites the *challenge from paucity*: How can Perry's few imaginary cases provide *sufficient* evidence for his *a posteriori* account, especially given that all his cases appear to belabour the same point? Here, it may seem difficult to avoid the impression that Perry's cases at best provide 'super-weak inductive support' (Cappelen and Dever 2013, 41) for his two-factor account of locating belief.

The fact that Perry exclusively relies on imaginary scenarios invites the *challenge from relevance*: How can the evidence from imaginary scenarios *be of the right sort* to support Perry's *a posteriori* two-factor account to begin with? As Cappelen and Dever convincingly insist, Perry cannot defend his exclusive reliance on imaginary cases by avowing: 'I am doing a form of psychology or cognitive science about very general truths about humans' (Cappelen and Dever 2013, 55). For '[t]here's no such thing as 'doing psychology about very general truths about human beings' *without empirical backing*' (ibid., my italics). I emphatically agree. Whatever else sound methodology requires, it patently requires us to provide empirical backing for any *a posteriori* hypothesis we accept on positive grounds.

Cappelen and Dever (2013, ch. 3, esp. 40, 55–56) enlist this second challenge to justify their necessity reading, which Perry rejects. Yet nowhere in his book-length rejoinder (2019) does Perry even discuss methodological issues.[3] So let us lend him a helping hand.

Meeting the challenge from paucity I: Perry's examples as puzzle cases

Gendler famously assumes that 'to perform a *thought experiment* is to reason about an imaginary scenario with the aim of confirming or disconfirming some hypothesis or theory' (Gendler 2004, 1154). By this standard, Perry's cases qualify as thought-experiments. However, even if we agree to class them so, we need to see that Perry's cases are thought-experiments of a rather specific kind. The way Perry uses his cases differs markedly from how Jackson (1982), Putnam (1975), and

Searle (1980) employ their respective cases of Mary, Twin Earth, and the Chinese Room.

Jackson, Putnam, and Searle employ their scenarios as arguments aimed at convincing the readers of specific proximal conclusions. Having told the tale of Mary, Jackson asks: 'What will happen when Mary is released from her black and white room [. . .]? Will she *learn* anything or not?' (Jackson 1982, 130). He answers this question himself: 'It seems just obvious that she will learn something about the world and our visual experience of it' (ibid.). This claim—that colour-wise physically omniscient Mary learns something when she sees something coloured for the first time—is the proximal conclusion Jackson aims at. It is not part of the story he relates, but rather a result arrived at on its basis. Jackson enlists this result as a premise in his anti-physicalist 'knowledge argument', of which his thought-experiment is but a first step.

Perry does not use his cases as arguments, and he does not draw proximal conclusions. We are told what the shopper, the professor, and the hiker think and do, and that is it. This difference in narrative structure reveals a difference in epistemic function, or so I contend. Taking a page from Cappelen (2012, see esp. ch. 9), we can bring out the difference this way: Scenarios such as Jackson's Mary aim to establish answers. This is what makes them argument-like. Perry's scenarios, on the other hand, are puzzle cases. They are meant to highlight questions. Put simply, Perry's tales say: 'Look here, this is what happens: $p_1, \ldots, p_n$ (where the *p*s are propositions capturing core aspects of Perry's cases). But *why* is it that $p_1, \ldots, p_n$? How are we to *explain* this?'

Perry's cases, thus, do not serve to establish proximal conclusions. Their epistemic function is rather to identify *explananda* by highlighting aspects that are, given the relevant background, in need of explanation. Here are two key *explananda* highlighted by the case of the shopper, the professor, and the hiker:

- Why does the behaviour of the agent change precisely when he becomes willing to assert the indexical statement '*I* am the shopper who is making a mess', 'The meeting starts *now*', or '*This* is the Mt. Tallac trail and *that* is Gilmore lake'?
- Why is it that we may no longer have an explanation of the agent's behaviour if we replace the indexical by a non-indexical designation of the person (or time, or location)?

In claiming that Perry's examples are puzzle cases, I do not suggest that they are exceptional. Puzzle cases are common in philosophy. Recall Russell's tale of George IV and Scott:

> If *a* is identical with *b*, whatever is true of the one is true of the other, and either may be substituted for the other in any proposition without

> altering the truth or falsehood of that proposition. Now George IV wished to know whether Scott was the author of *Waverley*; and in fact Scott *was* the author of *Waverley*. Hence we may substitute *Scott* for *the author of* 'Waverley', and thereby prove that George IV wished to know whether Scott was Scott. Yet an interest in the law of identity can hardly be attributed to the first gentleman of Europe.
> (Russell 1905, 485)

This story is not meant to establish a proximal conclusion. It rather poses a question we may approximate thus: How is it that George IV wished to know whether Scott was the author of 'Waverley', but that he did not wish to know whether Scott was Scott? Russell's tale is a puzzle case, too. The same holds true of Block's (1995, 233) tale of superblindsight, Armstrong's (1980, 59) story of the long-distance truck driver, and Dretske's (1988, 52–53) using a nickel to represent Kareem Abdul-Jabbar alluded to above.[4]

Puzzle cases come in varying degrees of fictionality. Some puzzle cases depict manifestly (science-)fictional situations. See, for example, most of Lewis's (1980) tales of prosthetic vision. Other puzzle cases come as reports of actual fact. Think, for example, of Fodor's tale of the intelligent behaviour of his cat (Fodor 1987, ix). Although Perry's cases depict events he has imagined, his cases are close to the factual end of this spectrum. This will prove important below (see above. 'The challenge from relevance: near-actual truths and empirical backing' below).

Meeting the challenge from paucity II: inference to the best explanation

Russell famously declares that '[a] logical theory may be tested by its capacity for dealing with puzzles' (Russell 1905, 484), and he presents the case of George IV and Scott as one of 'three puzzles which a theory as to denoting ought to be able to solve' (Russell 1905, 485). He goes on to argue that only his theory of descriptions meets this explanatory requirement; the competing theories of Frege and Meinong fail at this task, and moreover are beset by theoretical problems. A key part of Russell's argument for his theory of descriptions, thus, takes the form of an inference to the best explanation from *explananda* identified by puzzle cases.

I venture that Perry's argument follows the same template. Perry (2000, 27–29) uses the cases of the absent-minded shopper, the tardy professor, and the lost hiker to identify *explananda* $p_1, \ldots, p_n$ that, as Perry assumes, a theory of human locating belief has to account for. He then argues that single-factor theories cannot account for $p_1, \ldots, p_n$. Perry completes his abduction by arguing that his indexicalist two-factor account can.

The single-factor theories Perry (2000, 29–38) considers agree that a subject *S* has a locating belief just in case *S* stands in the *believing* relation to a proposition. Yet they disagree on the kind of proposition involved. Consider the proposition *p* the professor comes to believe when he is ready to assert 'The meeting starts now'. The first theory equates *p* with a Fregean thought, understood as a non-indexical descriptive content with an absolute truth-value. Suppose that *p* is the Fregean thought *that the meeting starts at noon on 12 February 1978*. This, however, is something the professor has believed all along. And even if we agree that he just acquired this belief, the theory cannot explain why this acquisition should prompt him to act when he does, or so Perry argues. The second theory equates *p* with a singular proposition containing the time the meeting starts as an element. Perry maintains that this theory cannot even explain the professor's change of belief, since he all along believed of the time the meeting starts that this is when the meeting starts. The third theory equates *p* with a relativised proposition modelled as a function from indices, understood as pairs of persons and times, to truth-values. Again, Perry argues that the professor can hold a belief with this content without being prompted to act.

Turning to his positive proposal, Perry (2000, 38–41) argues that his two-factor theory accounts for all the *explananda* identified. The professor's change in behaviour is not explained by a change in *what* he believes, i.e. the singular proposition attributing to a time that this is when the meeting starts. It is rather explained by a change in *how* he comes to think of this time. When the professor comes to think of this time as playing the *now*-role, he gets up and leaves. This explains why our explanation collapses if we get rid of the truly indexical element.

I will not pass judgment on Perry's assessment.[5] Rather, I will highlight two features of his methodology. First, the competing theories Perry discusses are generic approaches. Since these cannot account for general patterns in the data, Perry can rightfully claim that the explanatory inferiority he diagnoses projects to all single-factor representational theories, just as is required if his abduction is to show more than that his account is the 'the best of a bad lot' (van Fraassen 1989, 143; see ibid., 142–149; Lipton 2004, ch. 9).

Second, Russell's case of George IV and Scott identifies a single *explanandum*. Since Russell's theory is the only relevant competitor that can account for it, this *explanandum* provides rather strong inductive support for Russell's view. The analogous holds true for Perry. If Perry is right, his two-factor account of locating belief is the only relevant theory that can account for the *explananda* his cases identify. This explanatory success provides rather strong inductive support for Perry's view. When it comes to establishing a theory by an abduction, we find that a few well-chosen *explananda* can make all the evidential difference.

The challenge from relevance: near-actual truths and empirical backing

This leaves the hard part. We still need to understand why the *explananda* that Perry enlists are suited to empirically support his *a posteriori* account of locating belief in the first place. Perry accepts this account because it best explains $p_1, \ldots, p_n$. But we have no assurance that these propositions are actually true. They derive from a fictional narrative and for all we know, neither Perry himself nor anybody else ever actually followed a trail of sugar in a supermarket. But if $p_1, \ldots, p_n$ are not actually true, as I will acknowledge in order to avoid begging the question, how can the fact that it best explains these propositions provide any empirical support for Perry's *a posteriori* account? This support needs to be empirical. For, to repeat, whatever else sound methodology requires, it patently requires us to provide empirical backing for any *a posteriori* hypothesis we accept on positive grounds.

My answer to this question comes in two parts. Let me call a proposition p a *near-actual truth iff* p could easily be true, which is to say that p is true in some metaphysically possible world w close to the actual world.[6] Here, then, is the first part of my answer: The propositions $p_1, \ldots, p_n$ could easily be true, and we know this to be so. They are *known near-actual* truths, as I will say.

The propositions $p_1, \ldots, p_n$ depict perfectly ordinary sequences of events where a normal human being acts on her beliefs and desires whilst performing a mundane task such as shopping or hiking. Now, we know what people like you and me are prone to do, given what they believe and desire, when performing such mundane tasks. Our everyday dealings have taught us as much; there is no need for cutting-edge scientific psychology. As Perry puts it: 'everyone knows how people work' (Perry 2019, 73). Yet the way the shopper, professor, and hiker act fits the general pattern of what we quite generally know human agents to do; $p_1, \ldots, p_n$ cohere exceedingly well with our non-scientific body of knowledge about human agency. This remains true if we add any general background knowledge an educated armchair philosopher can draw on, including 'discoveries of specialised inquiry that were perhaps once cutting-edge enough to not be armchair, but now have achieved sufficiently common currency' (Nolan 2015, 215).

Suppose someone were to relate Perry's stories as known fact. Would we need to revise any of the beliefs we may draw on in the armchair? No. If we would choose to accommodate $p_1, \ldots, p_n$, we could simply add these propositions to our beliefs. Now if you know little zoology, you can easily accommodate the false claim that koala bears are bears. This is not what we find in Perry's cases. Our ordinary knowledge of human agency is comprehensive and detailed; so is arguably our armchair-available background knowledge. Judging by our ordinary knowledge

about human agency combined with armchair-available background knowledge, then, $p_1, \ldots, p_n$ may be true in the actual world. This does not show that these propositions are true. Yet it warrants us to hold that these propositions are true in some possible world *w* that is just like the actual world in all respects covered by that extensive body of knowledge. Any such world clearly qualifies as close to actuality. I conclude that we have every right to consider $p_1, \ldots, p_n$ to be known near-actual truths.

Let me forestall two objections to this line of thought. First: *How can you assume that we can reliably know that some proposition p is true in a nearby possible world without committing to a specific modal epistemology*? The answer is simple. All currently discussed modal epistemologies, including those proffered by self-professed 'modal skeptics' (Van Inwagen 1998; Hawke 2017), agree that we can reliably procure knowledge of mundane possibilities such as, say, that my new mug placed dangerously close to the edge of my desk *could* fall and break (see e.g. the contributions to Fischer and Leon 2017). They do so irrespective of whether they explain such modal knowledge by imagination-based counterfactual reasoning (Williamson 2007, ch. 5, 2015), by enlisting similarities between objects (Roca-Royes 2017), or by a projection from directly experienced modal properties in actions and affordances (Vetter 2020). Since whichever approach will eventually win out underwrites my assumption, I do not need to commit to any specific approach.

Let us turn to the second objection: *How can you take it for granted that we can know whether some p holds true in some nearby world(s) without committing to standards for assessing the similarity between worlds?* Here is my answer. We are not concerned with similarity for *all* propositions and *all* worlds. We are exclusively concerned with mundane propositions about human agency and worlds close to actuality. But when it comes to these propositions and those worlds, I hold that we may safely proceed on the assumption that a world *w* is close to actuality if *w* is governed by the same natural laws as actuality, if *w* quite generally exhibits overall the same regularities as actuality, and if almost all spatiotemporal regions of *w* show an overall match of particular fact to those of actuality such that *w* differs from actuality at most in some particular facts. This reflects Lewis's (1979b) original view. It also fits well with our intuitive convictions that, say, my new mug placed dangerously close to the edge of the desk could *easily* be swiped off, or that the meter rod could have *easily* been slightly longer than it actually is, as Kripke (1980, 55–56) argues.

This leaves the second step of my argument: Why should we agree that known near-actual truths provide empirical backing? The argument here is straightforward. Our empirical theories are measured by how well they account for what does happen *and* by how well they account for what *could easily* happen—if you like, their domains of application extend beyond the actual world to close possible worlds. Our theory of

falling bodies should not just specify the distance my old mug traveled within *n* seconds when it did fall. It should also specify the distance my new mug *would* travel within *n* seconds were it to fall. By the same token, our theory of human agency should not just explain human behaviour that did actually occur. It should also be capable of explaining human behaviour that could easily occur. We have already seen that Perry's two-factor theory meets this requirement. His Indexicality Lemma supports the counterfactual that *if* a human intentional agent *S* were to perform a Perry-type action *A*, then *S*'s motives for *A would* essentially comprise a truly indexical locating belief. The domain of application of Perry's theory, too, comprises the actual world, and those possible worlds close to it as well.

We have seen that known near-actual truths are a legitimate part of what we evaluate our theories by. If our theory of falling bodies gets wrong what would happen if my mug were to fall, as it easily could, this speaks against the theory. Conversely, a theory receives support when it accounts for events that could easily occur. I conclude that known near-actual truths are of the right sort to adequately support *a posteriori* hypotheses. Since $p_1, \ldots, p_n$ are known near-actual truths, these propositions are of the right sort to support Perry's account. The backing these propositions provide qualifies as empirical. They are empirical propositions, and our reasons to hold that they are near-actual truths are thoroughly empirical. How else would you want to characterise the evidential support the known near-actual truths $p_1, \ldots, p_n$ provide?

I have argued that mere near-actual truths can support empirical theories. This explains why we often do not care whether *explananda* are actual truths or merely near-actual truths. Suppose you learn that Scott did not write *Waverley* after all. So there never was any danger of George IV wishing to know whether Scott was Scott. Will you disregard this case when it comes to assessing Russell's theory of descriptions? You will not. Even if this is a mere near-actual truth, Russell's theory receives support by accounting for it. Or consider the following tale, that Ernst Mach relates:

> Not long ago, after a trying railway journey by night, when I was very tired, I got into an omnibus, just as another man appeared at the other end. 'What a shabby pedagogue that is, that has just entered', thought I. It was myself: opposite me hung a large mirror. The physiognomy of my class, accordingly, was better known to me than my own.
>
> (Mach 1914, 3 fn.1)

Mach reports this as a known actual truth. Would you treat Mach's story any different had he presented it as an imaginary case? I do not think you would. Any theory of identifying representations needs to

account for Mach's tale, regardless of whether it depicts historical fact, or merely a dreamt-up near-actual case.

A meta-philosophical upshot

Our in-depth case study of Perry's case brings home that an armchair philosopher engaging with *de facto* scientific hypotheses can legitimately exploit puzzle cases to procure empirical evidence that may allow her to pass judgment on competing empirical theories. I conclude that the 'puzzle-case abduction template' Perry follows is fundamentally methodologically sound. This marks an important meta-philosophical insight in its own right. But our in-depth case study affords meta-philosophical conclusions beyond it. I will focus on two that bear on Glock's meta-philosophical picture.

First: *The division of labour between armchair philosophy and the empirical sciences in the assessment of de facto scientific hypotheses is not as neat as Glock, Nolan, and Papineau have it*. It is not true that the empirical sciences provide all the relevant evidence, whereas philosophy merely enhances our theoretical or 'conceptual' understanding. Our in-depth case study shows that an armchair philosopher may well be in a position to procure relevant empirical evidence, too. This may work better when the empirical hypotheses under discussion are suitably general, and it may be easier in economics, semantics, or psychology than in physics, or the study of animal mind, for that matter. Still, our case study brings home that armchair philosophy is not just valuable because, when we assess scientific hypotheses, the armchair is a good position from which to theorise.

My second meta-philosophical conclusion runs directly counter to Glock's key axiom: *The proper domain of philosophy extends beyond conceptual issues*. Glock insists that '[t]heoretical philosophy is a *second-order discipline*. It does not directly describe or explain reality; instead it reflects on the concepts that we use outside philosophy, in everyday life, science, or other specialised domains' (Glock 2017, 80). Our in-depth case study says otherwise. Perry does not reflect on concepts. He rather devises a first-order theory of human locating belief aimed at explaining certain marked peculiarities in human intentional action. Nevertheless, Perry still does philosophy, or so it would seem.

Now Glock may want to insist that philosophy is, by the very concept, bound to be a second-order conceptual discipline. But why should this be more than a parochial stipulative restriction of the term 'philosophy?' Should we not rather look at the actual practice we find philosophers like Perry to engage in and conclude that philosophy does include first-order theorising about general contingent aspects of reality as well—which is what I have done?

Alternatively, Glock may want to insist that Perry deals in conceptual matters after all. In an intriguing twist to his 'impure analysis',

Glock attests 'mutual dependencies of conceptual and factual considerations' (Glock 2017, 97) and maintains that conceptual connections may depend on contingent facts. For instance, given that our closest evolutionary ancestor without language is not around any longer, '[h]aving a mind presupposes having a language' may amount to a conceptual truth. I agree that if we blur the distinction between conceptual analysis and general theorising in this way, Perry can rightly be said to attain conceptual insights. However, labelling Perry's results 'conceptual' rather than 'theoretical' would not change the fact that Perry attains his results by devising a first-order account apt to explain worldly goings-on, rather than by engaging in conceptual reflection. Any such blurring would moreover turn the claim that philosophy proper deals with conceptual issues into a triviality perfectly amenable to any philosophical naturalist averse to analytic truths.

Notes

1 Knobe's (2015) qualitative analysis finds that although the field is now dominated by empirically-based work, pure armchair argument is still common in current philosophy of mind. Ashton and Mizrahi (2018) identify a similar trend.
2 See Tye (1992) and Dennett (1988) for two papers that are particularly rich in imaginary cases.
3 The only explicit methodological claim I find is on p. 103, where Perry insists that his views about how chickens behave stem from experience, rather than from armchair reasoning.
4 Cappelen claims that philosophical cases exclusively serve to raise 'questions about some philosophically significant features of the world. (Contrast this with the idea that their function is to answer questions.)' (Cappelen 2012, 188). If he indeed thinks that all philosophical cases are puzzle cases, Jackson's Mary, Putnam's Twin Earth, and Searle's Chinese Room prove him wrong.
5 See Evans (1981) for a Fregean response to Perry. See Perry (2000, 1–26) for a reply, and Perry (2019) for an extended defence.
6 Kment (2014, 2–5, 30–34) prominently stresses the ubiquity and systematic importance of ordinary 'could easily'-locutions. I accept this *data*, but I do not accept the revisionist modal metaphysics that Kment argues for (see ibid., chs. 2–5).

References

Armstrong, D. (1980). *The Nature of Mind and Other Essays*. New York: Cornell University Press.

Ashton, Z. and Mizrahi, M. (2018). Show Me the Argument: Empirically Testing the Armchair Philosophy Picture. *Metaphilosophy* 49(1–2): 58–70.

Bach, T. (2019). In Defence of Armchair Expertise. *Theoria* 85(5): 350–382.

Block, N. (1995). On a Confusion about a Function of Consciousness. *Brain and Behavioral Sciences* 18(2): 227–247.

Cappelen, H. (2012). *Philosophy without Intuitions*. Oxford: Oxford University Press.

Cappelen, H. and Dever, J. (2013). *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*. Oxford: Oxford University Press.

Dennett, D.C. (1988/1993). Quining Qualia. In Marcel, A.J. and Bisiach, E. (eds.), *Consciousness in Contemporary Science*. Oxford: Clarendon Press, pp. 42–77.

Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. Cambridge: MIT Press.

Evans, G. (1981). Understanding Demonstratives. In his *Collected Papers*. Oxford: Clarendon Press, pp. 291–321.

Fischer, B. and Leon, F. (eds.) (2017). *Modal Epistemology after Rationalism*. Springer.

Fodor, J.A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Gendler, T.S. (2004). Thought Experiments Rethought—And Reperceived. *Philosophy of Science* 71(5): 1152–1163.

Glock, H. (2003). *Quine and Davidson on Language, Thought and Reality*. Cambridge: Cambridge University Press.

Glock, H. (2010). From Armchair to Reality? *Ratio* 23(3): 339–348.

Glock, H. (2013). Animal Minds: A Non-Representationalist Approach. *American Philosophical Quarterly* 50(3): 213–232.

Glock, H. (2017). Impure Conceptual Analysis. In D'Oro, G. and Overgaard, S. (eds.), *The Cambridge Companion to Philosophical Methodology*. Cambridge: Cambridge University Press, pp. 77–100.

Glock, H. (2019). Agency, Intelligence and Reasons in Animals. *Philosophy* 94(4): 645–671.

Hawke, P. (2017). Can Modal Skepticism Defeat Humean Skepticism? In Fischer, B. and Leon, F. (eds.), *Modal Epistemology after Rationalism*. Cham: Springer, pp. 281–308.

Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly* 32: 127–136.

Kment, B. (2014). *Modality & Explanatory Reasoning*. Oxford: Oxford University Press.

Knobe, J. (2015). Philosophers Are Doing Something Different Now: Quantitative Data. *Cognition* 135: 36–38.

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lewis, D. (1979a). Attitudes De Dicto and De Se. *The Philosophical Review* 88(5): 13–43.

Lewis, D. (1979b). Counterfactual Dependence and Time's Arrow. *Noûs* 13: 455–476.

Lewis, D. (1980/1986). Veridical Hallucination and Prosthetic Vision. In Lewis, D. (ed.), *Philosophical Papers*. Vol. II. Oxford: Oxford University Press, pp. 273–286.

Lipton, P. (2014). *Inference to the Best Explanation*. 2nd edition. London: Routledge.

Mach, E. (1914). *The Analysis of Sensation*. Translated by C.M. Williams and S. Waterlow. Chicago, IL/London: Open Court.

Nolan, D. (2015). The A Posteriori Armchair. *Australasian Journal of Philosophy* 93(2): 211–231.

Papineau, D. (2014). The Poverty of Conceptual Analysis. In Haug, M.C. (ed.), *Philosophical Methodology: The Armchair or the Laboratory?* London/New York, NY: Routledge, pp. 166–194.

Perry, J. (2000). *The Problem of the Essential Indexical and Other Essays*. Expanded edition. Oxford: Oxford University Press.

Perry, J. (2019). *The Essential Indexical Revisited*. Stanford: CSLI.

Putnam, H. (1975). The Meaning of 'Meaning'. In Putnam, H. (ed.), *Mind, Language, and Reality*. Philosophical Papers. Vol. 2. Cambridge: Cambridge University Press, pp. 215–271.

Roca-Royes, S. (2017). Similarity and Possibility: An Epistemology of de re Possibility for Concrete Entities. In Fischer, B. and Leon, F. (eds.), *Modal Epistemology after Rationalism*. Cham: Springer, pp. 221–245.

Russell, B. (1905). On Denoting. *Mind* 14(56): 479–493.

Searle, J. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences* 3(3): 417–457.

Tye, M. (1992). Visual Qualia and Visual Content. In Crane, T. (ed.), *The Contents of Experience*. Cambridge: Cambridge University Press, pp. 158–176.

van Fraassen, B. (1989). *Laws and Symmetry*. Oxford: Clarendon Press.

van Inwagen. (1998). Modal Epistemology. *Philosophical Studies* 92(1): 67–84.

Vetter, B. (2020). *Potential Knowledge*. Manuscript.

von Eckardt, B. (2012). The Representational Theory of Mind. In Frankish, K. and Ramsey, W. (eds.), *The Cambridge Handbook of Cognitive Science*. Cambridge: Cambridge University Press, pp. 29–49.

Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.

Williamson, T. (2015). Knowing by Imagining. In Kind, A. and Kung, P. (ed.), *Knowledge through Imagination*. Oxford: Oxford University Press, pp. 113–123.

8 Perception, causation, disjunction

John Hyman

I met Hanjo Glock in 1986, when we were both graduate students in Oxford, and I have learned from and admired his scholarship, his philosophical acumen, and his intellectual passion ever since. Contributing to Hanjo's festschrift is a perfect opportunity to express my gratitude for our friendship and for his contribution to philosophy. He is one of the relatively few philosophers today whose work is always alive to the historical roots of contemporary problems, and whose historical work illuminates philosophy today—above all, his work on twentieth-century analytical philosophy, which is where the topic of this essay belongs. It is as reminiscent—redolent even—of Oxford philosophy in the last century as any, the topic of perception.

British philosophy of perception conducted a step-by-step retreat from empiricism in the twentieth century, as Lockean indirect realism and phenomenalism gave way to the modern causal theory of perception advanced by Paul Grice and defended by Peter Strawson, and as that theory came under attack by Paul Snowdon, Mike Martin and others, under the banner of disjunctivism. I should make it clear right at the beginning that there are several versions of disjunctivism, which differ in significant ways (see Sturgeon 2008, 112–119). But I shall focus almost exclusively on Grice, Strawson, and Snowdon. Even so, I shall not attempt a comprehensive assessment of their views about perception. Instead, I shall begin with some introductory comments placing the modern causal theory against its empiricist background, and then examine Strawson's argument in favour of the theory and Snowdon's objection to it. I shall not contest the objection. On the contrary, I shall press it further, against the disjunctivism Snowdon himself defends. The burden of my argument is that the retreat from empiricism has not gone far enough.

The modern causal theory: conceptual analysis

In his last defence of the modern causal theory of perception, Strawson (1998, 311) states its cardinal claim as follows: 'it is a conceptual truth that when a subject S sees an external object O, O is causally responsible for S's visual experience'.[1] (The claim is meant to generalise across

DOI: 10.4324/9781003202929-11

the senses.) But there is more to the theory than this. I shall explain what more by setting out the arguments Grice and Strawson offer in its favour, and by exploring two key ways in which it differs from the classical causal theory advanced by Locke.

The principal argument supporting the theory is that the sensory experiences we have when we perceive features of the world around us match or correspond with these features to a high degree, although, of course, we are also susceptible to hallucinations, illusions, and mistakes. But—and this is the critical step in the argument—if the correspondence is coincidental, the experience will not qualify as a perception. For example, suppose I am barbecuing lamb chops in the garden, and my wife has the experience of smelling barbecued lamb chops. It would be a mistake to think that she is smelling *these* lamb chops, the ones on *our* barbecue, if the fact that we have lamb chops on our barbecue is purely coincidental and the delicious vapour she is inhaling comes from our neighbour's barbecue, and not our own. So, if the correspondence between our experience and our environment is coincidental, the experience will not qualify as a perception; and if it is *not* coincidental, then barring metaphysical fantasies such as occasionalism or pre-established harmony, the reason must be that our experience is responsive to its objects, in other words, that changes in the features of our environment that we perceive cause changes in our sensory experience to occur. Hence, the experiences that *do* qualify as perceptions must be caused by the objects we perceive.

Both Grice and Strawson argue along these lines. The principal difference between them is that they support what I called the critical step in the argument in different ways. Grice supports it by appealing to 'linguistic fact'—for example, the fact that we would not *say* that my wife was smelling the lamb chops on our barbecue in the circumstances described. Whereas Strawson supports it by pointing out that we think of perception as a source of knowledge, and belief (e.g., my wife's belief that there are lamb chops on our barbecue) does not qualify as knowledge if it is true fortuitously, or by chance. I shall examine both Grice's and Strawson's arguments in due course. But I shall begin with some preliminary remarks about the aim of the modern causal theory and its historical background.

I have called the theory of perception that Grice and Strawson defend the *modern* causal theory to distinguish it from the classical causal theory advanced by Locke. They differ in two principal ways. First, the modern theory is an exercise in conceptual analysis, whereas Locke means to provide a framework for scientific enquiry and a model of the mind. Second, it is a version of direct realism, or, more cautiously, it is not a version of indirect realism. I shall enlarge on these two points in turn.

First, in Grice's words (1961, 121–122), the modern theory is meant 'to elucidate or characterize the ordinary notion of perceiving a material

object'. On the one hand, it is not meant to define a concept that is better suited to scientific enquiry than the unscientific concept we learn to use as children. On the contrary, it is meant to elucidate or characterise that very concept. On the other hand, it is not sufficient, in order to count as accepting the theory, merely to hold that the perception of an object occurs as a result of a process that involves the object itself 'at an earlier stage'. 'Such a belief', Grice says (1961, 121), 'does not seem to be philosophical in character', because 'it has the appearance of being a very general contingent proposition'.

Similarly, without any commentary or context, Strawson's claim that the experiences enjoyed in sense perception are caused by the objects we perceive might look like 'a very general contingent proposition' or an elementary scientific fact. But he states emphatically in several places that this is not what he has in mind. For example, in *Analysis and Metaphysics*, he writes as follows:

> This notion of the causal dependence of the experience enjoyed in sense-perception on features of the spatio-temporal world [. . .] is not something we discover with the advance of science, or even by refined common observation [. . .]. It is conceptually inherent in a gross and obvious way in the very notion of sense perception as yielding true judgements about an objective spatio-temporal world.
> (Strawson 1992, 61)

And in his last published article on the topic, Strawson insists that the notion of causal dependence 'is integral to the ordinary concept of perception', and as already noted he re-states his key claim as follows (concentrating on the sense of sight): 'it is a conceptual truth that when a subject S sees an external object O, O is causally responsible for S's visual experience' (Strawson 1998, 311).

Why does this matter? Why should we care whether a belief is philosophical in character or whether a truth is conceptual? Suppose the objects we perceive do cause our perceptions. Do we need to decide whether the fact that they do so is implicit in the concept of perceiving an object or whether it is a very general item of scientific or proto-scientific knowledge, which has been known for hundreds or perhaps thousands of years? Besides, how sharp is the distinction between these kinds of truths? For example, is it a scientific truth or a conceptual truth that men beget children or that spaghetti does not grow on trees? Is this distinction between kinds of truths sustainable at all?

These sceptical questions cannot be brushed aside, but I shall assume here that the distinction between conceptual truths and scientific truths is legitimate.[2] One of the tasks of philosophy has always been to explain and clarify the main concepts or ideas in a domain of thought—and to criticise or modify concepts, when they give rise to paradoxes or embody

confusion. In some parts of philosophy, such as ethics or philosophy of perception as Grice and Strawson practice it, the domain of thought is one we all inhabit, simply in virtue of being mature, socialised human beings; in other parts of philosophy, such as philosophy of law or philosophy of perception as others practice it, it is not. I do not believe that the clarification of concepts is a futile exercise, or that it rests on a mistake. Be that as it may, it is important to understand that this is the task that Grice and Strawson set themselves, because it accounts for the ways in which they argue for the theory, in other words, the kinds of evidence they present in favour of it.[3]

Grice describes the evidence for the causal theory as 'linguistic fact', the fact that it would be correct or incorrect to say that someone saw a particular object, or the fact that we would be inclined to say that he saw it or did not see it, in a given set of circumstances. This would be a very peculiar way of arguing in favour of a scientific or proto-scientific claim about how perceptions occur, or, for that matter, in favour of the claim that men beget children or that spaghetti does not grow on trees. Imagine trying to disabuse a nineteenth-century Solomon Islander who doubts whether men play any role in procreation by analysing pillow-talk between parents, or a twenty-first-century Londoner who believes that spaghetti does grow on trees by examining table-talk in trattorie in Milan or Rome. But linguistic facts evidently are pertinent to a claim about the ordinary notion of perceiving a material object, because speech is an immediate manifestation of concepts—an activity in which the use of concepts can be directly observed.

As noted above, Strawson's argument in favour of the causal theory is different from Grice's, and it is not in any obvious or direct way about language. Here is another concise statement of it:

> It is certainly a feature of our ordinary scheme of thought that sense perception is taken to yield judgements which are generally or usually true. Remember that in thinking of the world as objective, we are thinking of it as being the way it is independently of any particular judgement about it; the truth of the judgement, if it is true, consists in its conformity to the way things are in the world. Hence the minimum that seems to be involved in the notion of sense perception generally yielding true judgements about an objective [. . .] world is that there should be some pretty regular relation of [causal] dependence of the experience enjoyed in sense perception on the way things objectively are.
>
> (Strawson 1992, 60)

If we set aside the question of whether this argument is convincing—we shall consider that question shortly—and focus only on its premises, we shall see that they are simply about relations between features of 'our

ordinary scheme of thought', in other words, relations between concepts or ideas. Such and such is 'a feature of our ordinary scheme of thought', Strawson says; and in thinking of A we are thinking of B. Again, these premises could hardly justify a scientific or proto-scientific conclusion about the operation of the senses, and this is clearly not what Strawson intends. His principal claim is that we cannot conceive of the senses as cognitive faculties, as faculties which enable us to learn about objects that exist independently of our experience, without implicitly thinking of our experience as responsive to, as modified by, these objects. It may be an exaggeration to describe this claim as 'gross and obvious', but perhaps we can confirm it by examining 'our ordinary scheme of thought', and its manifestation in our use of words. This is the method that Strawson and Grice employ.

The modern causal theory: direct realism

The modern causal theory of perception is an exercise in conceptual analysis. This is the first important way in which it differs from the classical causal theory advanced by Locke. The second is that Locke held (or is widely thought to have held) that the immediate objects of perception are mental images or ideas, which represent the physical objects that cause them, and in a sense stand proxy for these objects in our minds. Because of this doctrine, Locke's theory was commonly—although not universally—thought to imply that the existence of objects beyond the mind cannot be known, except perhaps as a result of making an inference from images or ideas to physical objects, which is hard to justify. Locke himself famously stated that 'the certainty of things existing *in rerum natura*, when we have the testimony of our senses for it, is not only as great as our frame can attain to, but as our condition needs' (Locke 1997, 9.11.8). Whether this is a judicious acknowledgement of human frailty and fallibility or a sign of insufficient intellectual rigour (i.e., a fudge) is a matter of judgement. Perhaps it is both.

Be that as it may, the doctrine that the immediate objects of perception are ideas or mental images of some kind remained prevalent among philosophers until the middle of the twentieth century, both among phenomenalists, such as the early Carnap and C.I. Lewis, and among philosophers who followed Locke, such as Quine and the early Ayer. It is criticised in *Sense and Sensibilia*, a brilliant and destructive series of lectures about perception, which Austin gave several times between 1947 and 1959. But Grice's article 'The Causal Theory of Perception' convinced many philosophers that it is possible to formulate a version of the causal theory that is not committed to this doctrine, and therefore also avoids the problem about knowledge, which made Locke's theory difficult to accept.

How did Grice manage this? He argued that if a person perceives a material object, the object is causally responsible for his sense-impression

or sense-datum, or (in a more cautious form of words) for 'a state of affairs' that is reported by 'some present-tense sense-datum statement' about him (Grice 1961, 152). But instead of introducing the terms 'sense-impression' and 'sense-datum' as names of a kind of object that is present in the mind of a person who is perceiving something, Grice introduced them contextually, by introducing the *sentence*, 'S has a sense-impression of O' as a paraphrase of a sentence, such as 'It seems to S as if he can perceive O', adding 'I shall myself [. . .] often for brevity's sake talk of sense-data or sense-impressions; but I shall hope that a more rigorous, if more cumbrous, mode of expression will always be readily available' (Grice 1961, 123–124).[4]

This way of introducing the terms 'sense-datum' and 'sense-impression' means that there is no need to prove 'the existence of objects of a special sort for which the term [. . .] is offered as a class-name' (Grice 1961, 123), because the term is not introduced as the name of a special sort of object, but—roughly speaking—as part of an abbreviation. And if a term is not the name of a special sort of object, then it is not the name of a mental object or an image, which represents a physical object in the mind. Thus, 'Tom has a visual sense-impression of Lucy' does not mean that there is a visual image of Lucy in Tom's mind. It simply means that it seems to Tom as if he sees Lucy. And so the claim that the experience of seeing Lucy is a visual sense-impression caused in a certain way by Lucy is perfectly consistent with the proposition that Tom can see Lucy directly, rather than by seeing an image of her in his mind. As Strawson puts it:

> We take ourselves to be immediately aware of real, enduring physical things in space [. . .]. The immediacy which common sense attributes to perceptual awareness is in no way inconsistent [. . .] with the causal dependence of [perceptual experience] on [the things we perceive].
>
> (Strawson 1979, 53)

Grice's manoeuvre was familiar by the time he used it. Bentham had made use of the same device, which he called 'definition by paraphrasis', to avoid postulating the existence of arcane legal objects, such as rights, and Russell had meted out a similar punishment to sets. In Grice's hands, a contextual definition transforms the impression caused by a physical object from an *object* of awareness into an *experience* of awareness. The private mental image is swept away, and with it—or so one hopes—the difficulty that Locke's theory faced, of explaining how knowledge can extend beyond mind. Grice's terminology is archaic: The psychological sense of the word 'impression' dates back to the seventeenth century, and it was introduced into philosophy by Hobbes. But it is a mistake to associate it in Grice's writings too closely with its forbears in the empiricist

tradition—with Hobbes's and Hume's 'impressions', Locke's 'ideas', or Russell's 'sense-data'. On the contrary, Grice's intention was to dispense with the arcane mental objects these terms were intended to refer to, and to retain, as he puts it, only the letter, but not the spirit of the classical causal theory. (Though I shall conclude that this is not an accurate description of the relationship between the two theories.)

The arguments

Now for the arguments. The argument I presented earlier about lamb chops is a non-sequitur, but I shall leave it as an exercise for the reader to work out why.[5] I shall now examine Grice's and Strawson's own arguments. First Grice:

> [I]t might be that it looked to me as if there were a certain sort of pillar in a certain direction at a certain distance, and there might actually be such a pillar in that place; but if, unknown to me, there were a mirror interposed between me and the pillar, which reflected a numerically different though similar pillar, it would certainly be incorrect to say that I saw the first pillar, and correct to say that I saw the second; and it is extremely tempting to explain this linguistic fact by saying that the first pillar was, and the second was not, causally irrelevant to the way things looked to me.
>
> (Grice 1961, 142)

We need to consider this argument in a couple of steps. First, the *immediate* reason why it would be incorrect (i.e., untrue) to say that Grice saw the first pillar and correct (i.e., true) to say that he saw the second is that in the set-up he describes he did *not* see the first pillar, and *did* see the second. Second, the simplest explanation of the fact that he did not see the first pillar is that it was hidden behind the mirror, and the simplest explanation of the fact that he did see the second pillar is that it was visible *in* the mirror.[6] And it seems perfectly possible to understand *these* explanations—banal and uninformative as they are, if one already knows the set-up—without postulating causal connections between the objects a person sees and the way things look to that person.[7] Why then does Grice regard the 'linguistic fact' he averts to as evidence for the causal theory? In effect, he relies on two theoretical claims: that the first pillar was hidden behind the mirror *because the mirror made it causally irrelevant to the way things looked to him*, and that the second pillar was visible in the mirror *because the mirror made it causally relevant to the way things looked to him*. But what is the justification for relying on them? Both claims are plausible, given some knowledge about optics, but Grice has not shown—or made any attempt to show—that they are implicit in 'the ordinary notion of perceiving a material object'. And if

we *assume* that they are implicit in it, this is tantamount to assuming that the causal theory is correct.[8]

In sum, semantic ascent does not serve a genuine purpose here. It creates the illusion that the argument reveals something about our concepts, but the 'linguistic fact' that Grice appeals to—the fact that it would be incorrect to say that he saw the first pillar and correct to say that he saw the second—is explained, trivially, in line with the principle that is (in)correct to say that *p* iff (not-)*p*. The fact that has a *non*-trivial explanation is a fact about what Grice sees in the set-up, not a fact about language, and absent an argument showing that the two theoretical claims are implicit in 'the ordinary notion of perceiving a material object', the explanation of *that* fact does not support the causal theory.

Like Grice, Strawson was attracted by the thought that if the correspondence between a person's experience and their environment is coincidental, the experience will not qualify as a perception. But he seems to have been dissatisfied with Grice's argument for the causal theory, and his own argument is rather different. It appears—with significant variations, as we shall see—in a number of publications, dating from 1974 to 1998. Here are the earliest and latest passages. First 1974:

> The concept of perception is too closely linked to that of knowledge for us to tolerate the idea of someone's being merely flukishly right in taking his M-experience to be the M-perception that it seems to be. Only those M-experiences which are in a certain sense dependable are to count as the M-perceptions they seem to be; and dependability in this sense entails dependence, causal [. . .] dependence on appropriate M-facts.
>
> (Strawson 1974, 71)

(By an M-experience, Strawson means the kind of experience one is having if it seems to one as if one can see or hear or in general perceive an object of some kind. Grice (1961, 121–122) uses the terms 'impression', 'sense-impression', and 'sense-datum' in the same way. For example, as you read this sentence it seems to you as if you can see a page of text. By an M-perception, Strawson means an instance of someone's actually perceiving something—for example, your actually seeing a page of text. And by an M-fact, he means the fact that an object of some kind, such as a page of text, exists.)

Now 1998:

> In order for an experience to amount to a genuine perception of an object (and hence a way of gaining knowledge about it) there must be such a relation between object and experience as to rule out the case of a subject's being merely flukishly or accidentally right in taking it that there is just the object before him that he takes himself to

> be perceiving [. . .] the relation of causal dependence [. . .] remains the only plausible candidate [for this relation].
>
> (Strawson 1998, 314)

The arguments in these two passages are similar, but not as similar as they may look at first glance. One difference is the retreat from 'entails' to 'remains the only plausible candidate'. But the difference I want to focus on is that in the 1974 passage Strawson is explaining what he thinks it takes for an M-experience 'to be the M-perception that it seems to be'. He is evidently assuming that an M-perception is a specific kind of M-experience (or sense-impression, in Grice's terminology), and explaining what distinguishes this kind of M-experience from the rest. In the 1998 passage, by contrast, he carefully avoids making this assumption, or at least making it explicitly—for reasons I shall explain in due course. But why does the difference matter? It matters because if we make the assumption, then we are naturally going to ask the question that Strawson does ask, namely, what kind of M-experience (or sense-impression) *is* the perception that it seems to be? And since a perception does not carry a hallmark, a sign of authenticity stamped on it, the answer will have to refer to a *relation* in which an M-experience can stand to something else—presumably the object of which it seems to be a perception, unless, as in some idealist and occasionalist systems, it is God. And this takes us a good part of the way towards endorsing the causal theory's principal claim. The only remaining step is to claim, as Strawson does, that 'causal dependence' is the only relation capable of explaining how an M-experience can be, as he puts it, 'in a certain sense dependable', that is, a dependable source of beliefs, and thereby a source of knowledge, about the world. So, the assumption provides a fast track to the causal theory. But is it true?

Are perceptions sense-impressions?

It should be obvious, and is not disputed by philosophers who hold that perceptions are sense-impressions, that *X* may perceive *Y*, whether it seems to *X* as if he perceives *Y* or not. For example, if Tom sees Lucy fleetingly in the distance, it *may* seem to Tom as if he sees Lucy, but it may instead seem to him as if he sees Caroline, or as if he sees somebody, but nobody in particular. If a perception were a kind of sense-impression, Tom's perception would be a different kind of sense-impression in each case: respectively, a sense-impression of Lucy, a sense-impression of Caroline, and a sense-impression of somebody, but nobody in particular. But it would be a perception of Lucy in every case.

What is less obvious is that *X* may perceive *Y* without having a sense-impression that *could* be the perception, because the perception is subliminal or unconscious. This is less obvious precisely because we are not

conscious of unconscious perceptions, but it had been well known for several decades by the 1960s, and it is surprising that neither Grice nor Strawson considers the bearing of unconscious perception on the causal theory.[9] Perhaps the Cartesian equation of thought and consciousness still exerted a subliminal influence on philosophers who assumed that every perception is a sense-impression, and hence that it is not possible to perceive an object unless it seems to one as if one is perceiving *something* (even if it is merely a shadow or a fleck on the horizon), which, whether one realises it or not, *is* the object in question. Perhaps they were influenced by the fact that one would not normally *say*, or be entitled to say, that one saw something, if the perception was unconscious; although the question of whether one would say or be entitled to say something must not be confused with the question of whether it could be true. This seems to have influenced Strawson's thinking in particular, because he devises the form of words that he considers apt to describe a sense-impression by asking how someone could formulate a description of their 'current visual experience' that 'confine[d] itself strictly within the limits of the subjective episode' by modifying a spontaneous description of what they were actually seeing. (In effect, they are supposed to replace 'I see [. . .] ' with 'It sensibly seems to me just as if I were seeing [. . .]') But, of course, a person's spontaneous description of what they are seeing—such as Strawson's example (1979, 97), gazing across the grounds of Magdalen College before the ravages of Dutch Elm disease, 'I see the red light of the setting sun filtering through the black and thickly clustered branches of the elms [. . .]' etc.—is inevitably confined to what they are *consciously* seeing, and so it excludes unconscious perception *a priori*.

Be that as it may, unconscious perception has been the subject of numerous empirical studies since the last quarter of the nineteenth century, including an early article by Peirce and Jastrow, and the examples are legion. The theory of unconscious perception is a lively area of debate, but the reality of the phenomenon is not in doubt.[10] And although the best-known examples are pathological—e.g., blindsight, which was not discovered until the 1970s—studies of perception in the periphery of the visual field suggest that it is a perfectly ordinary phenomenon, by no means limited to subjects with brain lesions.

Strawson's assumption in the 1974 passage that a perception is a kind of sense perception is, therefore, mistaken. It might be objected that unconscious perception is a recent discovery, and therefore not 'integral to the ordinary concept of perception'. But I am not suggesting that the existence of unconscious perception has always been known, only that it is not conceptually inherent in a gross and obvious—or refined and unobvious—way in the very notion of sense perception that it *cannot* be unconscious. Heliocentrism is a fairly recent discovery, but it is not ruled out *a priori* by 'the ordinary concept' of the sun. Wave-particle duality

was only discovered a century ago, but it is not ruled out *a priori* by 'the ordinary concept' of light. Equally, unconscious perception is not ruled out *a priori* by 'the ordinary concept of perception', however strange it may seem to those of us who imbibed empiricist orthodoxies as students. Otherwise, it would not be a discovery, it would be a contradiction in terms.

Returning to the terminology of the 1974 passage, Strawson's claim that dependable M-experiences 'count as the M-perceptions they seem to be' is mistaken. But, in fact, there are two mistakes here rolled into one: first, the assumption that a perception is a kind of M-experience; and second, the idea that M-experiences seem to be perceptions.

Strawson's position is that Tom's sense-impression of Lucy—in other words, its seeming to Tom as if he perceives Lucy—seems to Tom to be a perception of Lucy, and the causal theory identifies a condition it needs to satisfy in order to *be* the perception that it *seems* to be. But this cannot be right. For 'its seeming to Tom as if he perceives Lucy' is a sentence-nominal—a noun-phrase derived from a sentence—which refers to the state of affairs described by the sentence from which it is derived: 'It seems to Tom as if he perceives Lucy'. The sentence-nominal is produced in a standard way, by replacing the main verb of the sentence with a gerund and putting the subject in the genitive case. 'It is hot in Naples' yields 'its being hot in Naples' in the same way, and 'Brutus killed Caesar' yields the so-called imperfect nominal 'Brutus's killing Caesar' and the perfect nominal 'Brutus's killing of Caesar'.[11] But there is no more truth in the idea that its seeming to Tom as if he perceives Lucy is something that seems than there is in the idea that its being hot in Naples is something hot or that Brutus's killing of Caesar is something that kills Caesar. Tom's sense-impression of Lucy does not *seem* to be a perception. By definition, it is a seem*ing*, not a thing that seems.

Since its seeming to Tom as if he perceives Lucy is its seeming to Tom as if he stands in a particular relation to Lucy—the relation expressed by the verb 'sees' in the sentence 'Tom sees Lucy'—it is instructive to compare a sense-impression with a different case of seeming to stand in a relation. For example, suppose it seems to Tom as if he is married to Lucy: colloquially, Tom feels married to Lucy. Perhaps he began to feel married a few months after they were married, or perhaps he felt married as soon as the ceremony had ended, or perhaps he feels married despite the fact that they are not actually married, for example, because they are cohabiting, or because their marriage was not validly contracted. In any case, suppose Tom feels married to Lucy. Tom's feeling married to Lucy does not feel like a marriage, whether it is a true impression or a false one, and however it was caused. It may have been caused 'in the right way' by an official licensed to perform a marriage ceremony. No matter. Whatever its origin or cause, Tom's feeling married to Lucy neither *feels as if* it is a marriage nor can it ever *be* a marriage, even if it is planted in his

mind by God. It is not an experience that feels like a marriage, or seems to be a marriage, and the thought that only those feelings-like-marriage or seemings-to-be-married that are in a certain sense dependable are to count as the marriages they feel like or seem to be is obviously confused. One relation can feel or seem like another relation—cohabitation can feel like marriage—but feeling as if one is married is not a relation, and cannot seem as if it is one. A person's feeling married is his feeling as he would feel, or as he imagines he would feel, if he were conscious—not merely cognizant—of his married state. Similarly, its seeming to Tom as if he sees Lucy is its seeming to Tom as it would seem, or as he imagines it would seem, if he consciously saw Lucy. So, the concepts of being married and feeling married are evidently related to each other, in a way that it is not difficult to explain, and the same is true of the concepts of perceiving something and its seeming to one as if one perceives something.

Strawson's claim that 'only those M-experiences which are in a certain sense dependable are to count as the M-perceptions they seem to be' is doubly misconceived. For M-experiences neither are nor seem to be perceptions. The claim is, in fact, a residue of the classical causal theory of perception. Remember, according to the classical theory, the immediate object of perceptual awareness is an image, which represents a physical object in the mind. Only the subject's awareness of a dependable image is to count as a perception, and a dependable image is one with the right kind of cause. The modern theory was meant to eliminate the mental object of awareness, but in reality it transferred part of the classical conception of a mental image to the sense-impression. For the impression resembles the classical theory's *awareness of an image* in being an experience, and a candidate for the title of a perception, but it resembles *the image itself* in being dependable if caused in the right way.

In sum, if we assume that some sense-impressions are perceptions, we are bound to ask what distinguishes them from the rest. The distinguishing feature cannot be an intrinsic property of these sense-impressions themselves—they do not have a hallmark stamped on them, which guarantees their authenticity—so their status as perceptions must depend on a relation in which they stand to something else. It is, therefore, plausible to hold that the authentic sense-impressions, like photographs, are the ones that are caused (in the right way) by the objects they are sense-impressions of—regardless of whether a sense-impression is conceived of as an object of awareness or as an experience of awareness. But the initial assumption is mistaken. Perceiving is not an elite suburb of seeming-to-perceive, any more than knowing is an élite suburb of believing, *a fortiori* it is not seeming-to-perceive with a specific kind of cause.[12]

If this is right, the conception of sense perception that underlies the modern causal theory is an unstable compromise between empiricist representationalism and direct realism, as the latter is now generally understood, that is, the doctrine that a sense perception is an irreducible

relation between the subject and the object of perception—no less so than a marriage—in other words, as Mike Martin puts it (1997, 85), a 'relational state of affairs'. In his influential book *Radical Embodied Cognition*, Anthony Chemero writes:

> When an animal perceives something directly, the animal is in non-mediated contact with that thing. This implies, of course, that the perceiving isn't inside the animal, but rather is part of a system that includes both the animal and the perceived object.
>
> (Chemero 2009, 5.3)

The 'this implies, of course' is what Strawson denies. The argument above suggests that while the 'of course' is an exaggeration, the 'this implies' is true.

As far as I know, Strawson only commented once on the idea that sense perception is an irreducible relation—'a relational state of affairs'—and even here he addressed the idea obliquely. His comment is surprisingly dismissive: 'Only someone temporarily blinded by philosophy', he writes, 'could dream of denying that when a subject sees an external object, the visual experience enjoyed by the subject is one thing or occurrence in nature and the object seen is another and distinct thing in nature' (Strawson 1998, 314). But considered as an objection to the relational idea, this is not merely dismissive, it is wrong. For if Tom's perception of Lucy is a relational state of affairs, in which Tom and Lucy are the relata, it does not follow that Tom's perception and Lucy are one and the same thing or occurrence in nature. On the contrary, it follows that they are distinct things, since an instance of a relation cannot be one of its own relata. For example, a marriage is not the same thing as a spouse. (In Strawson's terminology (2000, 46), Lucy and Tom are substantial particulars, whereas their marriage is a non-substantial substance-dependent particular, as is Tom's perception of Lucy.)

Unlike Strawson, David Armstrong engages with the relational idea directly:

> There is a most serious objection to [the] attempt to construe mental states as relations to things in the world. Suppose, as is perfectly imaginable, that I have exactly the same perceptual experience as I had when I looked at the tree, but suppose that this time there is no tree there [. . .] [In this case] there is, by hypothesis, nothing in the world for me to be 'mentally related' to. So no unique, irreducible, relation can be involved. Yet, also by hypothesis, the mental state is no different from the mental state in [the case where I looked at the tree]. So no mental relation of ourselves to things in the world is ever involved.
>
> (Armstrong 1993, 39)

But this argument fails for a different reason. In this passage, 'perceptual experience' needs to read as meaning sense-impression, since what is perfectly imaginable is that one should have exactly the same sense-impression when there is no tree present. By hypothesis, the sense-impression—its seeming to Armstrong that he sees a certain kind of tree, from a certain point view, in a certain light—is no different in the two cases. But the mental states that are said to be relations to things in the world are not sense-impressions, they are perceptions. Armstrong simply assumes that a perception is a kind of sense-impression, and therefore finds it natural to infer from the fact that sense-impressions are not 'relations to things in the world' that perceptions are not either.

Like Strawson, Armstrong fails to disprove the relational idea, but his argument confirms its incompatibility with the assumption that a perception is a kind of sense-impression, which Strawson and Armstrong, whose theories of perception differ in other ways, both make (Armstrong 1993, 236–237). If we reject this assumption, as I have argued we should, we need a theory of perception that is compatible with the existence of unconscious perception, and which breaks with the empiricist tradition more completely than the modern causal theory of perception does. Describing perception as an irreducible relation is the right first step, but it does not get us very far—any more than it would if we were theorising about marriage. I shall outline a theory that I favour in the final part of this chapter, but first I shall look at Snowdon's objection to the modern causal theory and Strawson's 1998 passage.

Snowdon's disjunctivism

I have argued that a perception is not a kind of sense-impression, *a fortiori* it is not a sense-impression with a particular kind of cause. Snowdon challenged the causal theory in a series of articles published between 1981 and 2011, on partly similar grounds. Snowdon accepts that statements such as, 'It seems to X as if he can see an oasis' or 'It looks to X as if there is an oasis in front of him' may be true whether X is actually seeing an oasis or experiencing an hallucination. But he argues that the reason for this is not that perception and hallucination are experiences of the same kind with two different kinds of cause: The reason is that despite their syntax, 'looks' and 'seems' sentences should be construed as disjunctions, whose disjuncts are (as he puts it) made true by two quite different kinds of states of affairs. For example, 'Tom is married' is true if and only if Tom is married, whereas 'Either Tom is married or he is insane' is true whether Tom is married or insane, although marriage and insanity are (as most people will admit) different states of affairs. According to Snowdon, something similar is true of the kind of statement that Strawson regards as a description of a 'slice of sensible

experience'. The disjunctivist picture, he says, divides what makes this kind of statement true into two classes:

> In cases where there is no [visual perception] they are made true by a state of affairs intrinsically independent of surrounding objects; but in cases of [visual perceptions] the truth-conferring state of affairs involves the surrounding objects.
>
> (Snowdon 1980, 186)

In sum, according to Strawson, the statement that it seems to X as if he can see an oasis in front of him, or (the form of words that Snowdon discusses in these passages) the statement that it looks to X as if there is an oasis in front of him is true, if it is true, because X is having an experience that may or may not qualify as a perception, depending on how it was caused. However, Snowdon claims that it is true 'in virtue of two distinct sorts of states of affairs':

> either there is an object which looks to be an oasis to *X* (this is the case where an object is seen), or it is to *X* as if there is something of that sort happening (*X* is hallucinating an oasis). It is allowed, according to this, that the two cases which are described in the same way [. . .] might be of a quite different nature.
>
> (Snowdon 1990, 129)[13]

Snowdon's disjunctivist idea is, I submit, partly right and partly wrong. Snowdon is right in thinking that perceptions and hallucinations are 'of a quite different nature'. For *X*'s perception is an instance of a relation—between *X* and the object he perceives—whereas *X*'s hallucination is not. As Snowdon puts it, an hallucination is 'intrinsically independent of surrounding objects'. But he is wrong in thinking that the claim that it looks to *X* as if there is an oasis in front of him describes, or is made true by, two different sorts of states of affairs, one 'intrinsically independent of surrounding objects' and the other not. In the case where *X* is actually seeing something, the state of affairs that makes the claim that it looks to *X* as if there is an oasis in front of him true is the same state of affairs as the one that does so when he hallucinates.

Recall for a moment the example of Tom's feeling married to Lucy. Disjunctivism about 'feels-ascriptions', such as 'Tom feels married to Lucy', would combine the idea that the illusion of being married and actually being married are 'of a quite different nature', which is true, with the idea that 'Tom feels married to Lucy' describes, or is made true by, both sorts of states of affairs, which is false. 'Tom feels married to Lucy' never describes, and is never made true by, a marriage. It can only ever describe or be made true by a feeling—regardless of Tom's marital status, and regardless of how the feeling it describes, or is made true

by, is caused. Perhaps the confusion is easier in the case of perception, because both hallucinations and perceptions are mental states, whereas feeling married is and being married is not a mental state. But the logical point is just the same. Feelings, not marriages, make 'feels-ascriptions' true, and sense-impressions or M-experiences, not perceptions, make 'looks-ascriptions' true, whether or not they are veridical feelings or impressions, and however they are caused.

Returning to Snowdon's own example, if *X* sees *Y*, and *Y* looks to him like an oasis, (i) *X's perception of Y*, and (ii) *X's sense-impression of an oasis* are themselves two distinct sorts of experiences of a quite different nature. They are distinct, since either can occur without the other occurring,[14] and although we can call them both 'visual experiences' or 'mental states' if we wish, they are of a quite different nature, because (i) is an instance of a relation whereas (ii) is not. When *X* sees something that looks to him like an oasis, both states of affair obtain; when *X* hallucinates, an oasis only (ii) does. But whether *X* is perceiving or hallucinating, the statement that it looks to *X* as if there is an oasis in front of him is made true, if it is true, only by (ii).

Why does Snowdon embrace the ostensibly implausible idea that despite their syntax 'looks' and 'seems' sentences, such as 'It looks to *X* as if there is an oasis in front of him', should be interpreted as disjunctions? The reason, I suggest, is that he has not entirely freed himself from the way in which Grice and Strawson conceive of perceptual experience. For he accepts the (false) assumption that Grice and Strawson make that 'It looks to *X* as if there is an oasis in front of him' can describe or be made true by either a perception or an hallucination, despite (rightly) insisting, against them, that these are two distinct sorts of experiences 'of a quite different nature', and not one sort of experience with two different kinds of cause. This is the uncomfortable combination of ideas which makes it appear as if its looking to *X* as if there is an oasis in front of him cannot be a unitary state of affairs. But the assumption is a mistake. 'It looks to *X* as if there is an oasis in front of him' cannot describe or be made true by a perception. It always describes a sense-impression—whether *X* is seeing something, such as an oasis or a mirage, or not—and so it always describes the same kind of experience or mental state.[15]

In sum, Snowdon criticises Strawson's assumption that, considered in themselves, independently of their causes and effects, perceptions and hallucinations are essentially the same kind of experience, the kind of experience we can describe by means of 'seems' or 'looks' sentences. According to Snowdon, the assumption is either false or unproven. False, because in fact perceptions and hallucinations are essentially different kinds of experience, and 'seems' or 'looks' sentences should be construed as disjunctions, which describe, or are made true by, experiences of both kinds. Or unproven, because it has not been shown that this is not the case. But the objection is partly right and partly wrong: partly right

because perceptions and hallucinations are essentially different kinds of experience; partly wrong because 'looks' and 'seems' sentences do not describe perceptions.

The 1998 passage

I shall turn now to Strawson's argument in the 1998 passage. Remember, the important difference between this passage and the 1974 one is that in the later passage Strawson tries to formulate his argument for the causal theory in a way that avoids reliance on the assumption Snowdon had contested, that a perception is a specific kind of sense-impression. Here is the passage again:

> In order for an experience to amount to a genuine perception of an object (and hence a way of gaining knowledge about it) there must be such a relation between object and experience as to rule out the case of a subject's being merely flukishly or accidentally right in taking it that there is just the object before him that he takes himself to be perceiving [. . .] the relation of causal dependence [. . .] remains the only plausible candidate [for this relation].

How does the argument fare, without the assumption? The answer is that it fails, because what the long first sentence says is false: In order for an experience to amount to a genuine perception of an object, it does *not* have to be related to the object in a way that prevents the subject's belief from being merely flukishly right. For example, suppose you are watching the competitors as they warm up for a race. You rightly take yourself to be seeing the famous sprinter Sally Fleetfoot at the blocks, but you are unaware that Sally is impersonating her twin sister, who is also a sprinter and was supposed to be competing in this race. In these circumstances, you are, indeed, merely flukishly or accidentally right in taking it that there is just the person before you that you take yourself to be perceiving. Hence, however your experience and its object need to be related in order for you to perceive it, the relationship does not rule this out.

The literature contains many examples that prove this point, that is, prove that one *can* be merely flukishly right in taking it that there is just the individual or kind of object before one that one takes oneself to be perceiving. For instance, in Alvin Goldman's well-known story about papier-mâché barn façades, Henry, who is driving in the countryside with his son, points out what he rightly takes to be a barn:

> Unknown to Henry, the district he has entered is full of papier-mâché facsimiles of barns. These facsimiles look from the road exactly like barns, but are really just façades, without back walls or

> interiors, quite incapable of being used as barns. They are so cleverly constructed that travelers invariably mistake them for barns. Having just entered the district, Henry has not encountered any facsimiles; the object he sees is a genuine barn. But if the object on that site were a facsimile, Henry would mistake it for a barn.
>
> (Goldman 1976, 773)

Goldman is interested in explaining why, as he puts it, 'we would be strongly inclined to withdraw the claim that Henry *knows* the object is a barn'. But setting this question aside, indeed setting aside the question of whether we would, or would always, be inclined to withdraw the claim, Henry is certainly merely flukishly or accidentally right in taking it that there is a barn before him. So, the relation between the barn and Henry's experience of seeing it, whatever exactly it is, does not rule this out. But Henry's experience *is* a genuine visual perception of a barn.

The argument in the 1998 passage fails, because genuine perception does not exclude a flukishly right judgement about the individual or kind of object being perceived. The concept of sense perception is 'closely linked' to that of knowledge, as Strawson says. But while the perception of objects in our environment is a source of factual knowledge, it does not guarantee it, and so luck-excluding conditions that apply to knowledge do not transfer to perception. (Perceiving *that something is the case* is a different matter. For example, seeing *that the thing one is looking at is a barn* is a case of knowing a fact. So any luck-excluding condition that applies to factual knowledge generally applies here too.)

Escaping empiricism

How could Strawson have missed this obvious point? I suspect the reason is that he was still gripped by the fallacy that a perception is a kind of sense-impression, which, as we have seen, makes the argument for the causal theory difficult to resist. Be that as it may, in the final part of this chapter, I shall comment briefly on the question of how plausible Strawson's claim remains—the claim that a causal connection between object and experience is 'integral to the ordinary concept of perception'—once we have rejected the fallacy. And then, finally, I shall sketch an approach to the theory of perception that is free, or at least freer, from the empiricist model, which neither Strawson nor Snowdon was quite able to discard.

Concerning the question, it is uncontroversial that we could not perceive physical objects if they did not cause changes to occur in our bodies—directly, in the photoreceptors and the other peripheral neurons on which our senses depend, and indirectly, in the parts of the brain involved in sense perception. Interestingly, however, vision admits an exception to this rule. It is true that we cannot taste a substance that

does not affect the chemo-receptors in our tongues and noses, and we cannot hear a vibration that does not affect the auditory nerve. But matt black objects do not emit or reflect any light, or sufficient light to affect the photoreceptors in our retinas, and yet they are visible, unless their background or surroundings are black too; and if one matt black object exactly masks another, its presence does not make a difference to the pattern of excitation caused by the visible scene of which it is a part. This disproves the simplistic idea that every object we perceive initiates a sequence of changes in the body, which terminates in the perception. If we envisage something of this kind, we shall have to treat a visible scene holistically, and not imagine that each individual object in the scene causes a discrete part of the experience that occurs when we perceive it.

However, we can still ask whether the fact that the physical objects we perceive cause our perceptions—either individually or collectively—was discovered when the scientific attitude to nature and to human life began to develop in the ancient world, or whether it is integral to the ordinary concept of perception, and therefore part of what Strawson famously described as 'a massive core of human thinking which has no history—or none recorded in histories of thought' (Strawson 1959, 10).

Prosaic at it sounds beside this resonant phrase, I favour the first answer. The most we can say is that sense perception cannot be explained scientifically, unless we assume that the objects we perceive cause changes in our bodies—directly, in our sense organs, and indirectly in the other parts of the body involved in sense perception. Certainly, by the seventeenth century, science had progressed far enough for it to be obvious to any informed person who was free from the influence of metaphysical fantasies such as occasionalism, that perception would be utterly mysterious if this were not the case. But the concept of perception is not the concept of a mental state with a special kind of cause, and sentences in which perceptual verbs occur are perfectly intelligible independently of the causal idea. Even the thought that the causal idea must be accepted as a precondition for any scientific study of perception flies in the face of the historical facts. For geometrical optics progressed independently of the study of the physics and physiology of vision at least until the eleventh century, when all three parts of optics were integrated into a single comprehensive theory by the Arab scientist Ibn al-Haytham (known as Alhazen in the West).[16]

Once we reject the assumption that a perception is a kind of sense-impression, the claim that the causal connection between object and experience is 'integral to the ordinary concept of perception' loses credibility, and it is no longer difficult to see that the causal theory is in reality a proto-scientific *picture* or *model* projected onto 'our ordinary scheme of thought': The object causes changes in our sense organs, and these cause sense-impressions in our mind. The confused idea that a perception is a kind of sense-impression seemed to license an *a priori* argument

in favour of this model, but once we have set that idea aside, it does not take a profound study of history to understand that the first step—the object causes changes in our sense organs—was a contested theoretical claim until Alhazen's synthesis became the accepted framework for research in optics, and the second step—changes in our sense organs cause sense-impressions in our mind—is a muddled philosophical doctrine, which was established as an orthodoxy by Locke and modified, but not really abandoned, by Grice. Grice concludes his article with the suggestion that his version of the causal theory, 'however close to the letter, is very far from the spirit of the original theory' (1961, 152). Sixty years on, this seems the reverse of the truth: However far from the letter, it is close to the spirit of the original theory.

So, the claim that a causal connection between object and experience is 'integral to the ordinary concept of perception' must be rejected. (Remember, the 'ordinary' concept is not one that only scientifically literate or well educated people can possess: It is the unscientific concept we learn to use as children, and share with *everyone* who understands that 'I saw (heard, tasted, etc.) it' can be an answer to the question 'How do you know?'—the principle Strawson relies on—and knows which pillar Grice sees in the set-up he describes.) But rejecting the claim is not enough. We need to make a more radical break with the modern causal theory than this. It is true, as Strawson says, that sense perception enables us to form true judgements about the world. But this should not be our starting point when we reflect on the ordinary concept of perception, because infants must learn to perceive objects in their environment before learning to make judgements, and because most sentient animals cannot learn to form judgements. And it is true that learning to form true judgements about the world depends on *conscious* perception. But this should not be our starting point either, because animals that experience conscious sense perception experience unconscious sense perception as well, and because it is arguable—although not certain—that arthropods perceive objects despite not having consciousness at all (see Feinberg and Mallatt 2017). For both these reasons, it is helpful to start with unconscious perception—or better, perception as such, without the peculiar quality of being conscious—and then consider what is special about conscious perception, once we have a tolerably clear understanding of unconscious perception, or perception as such.

So, what is perception as such? The answer, I suggest, is that it is a multi-track disposition, a particular kind of responsiveness to the objects and properties perceived. I have argued elsewhere that knowledge of a fact is an ability—not the ability to perform a specific kind of act or to engage in a specific activity or range of activities, but rather a multi-track disposition, in Ryle's sense of the term (see Ryle 2009, 32). It is the ability to be guided by a fact, to respond to it rationally, in what one thinks, or feels, or does.[17] Cognition of things is similar to (not the same

as) cognition of facts. Think of a cat stalking a bird. The bird hops this way, the cat turns this way; the bird flutters across the courtyard, the cat advances a few paces; and so on. The cat's movements are responsive to, are guided by, the bird. Or think of a hiker following a guide. The guide takes the left path, so the traveller takes the left path; the guide pauses, so the hiker pauses; and so on. Whether one is guided by facts or by things, one is responsive to what one is guided by. But as Wittgenstein (PI, §§ 156–173) pointed out, this is not like a train being guided by the rails. For sense perception—like factual knowledge—is highly plastic, and how it gets expressed depends on one's purposes or goals. Furthermore, an animal's capacity for goal-directed behaviour encompasses controlling the operation of its own sense organs—by changing their orientation, by approaching, touching, or retreating from objects, etc.—and this control over sense perception, which can be conscious or unconscious, like perception itself, is an essential part of the normal life of every animal that is capable of goal-directed behaviour of any kind at all.

All of the highly varied ways in which sentient animals feed, mate, navigate, and communicate involve this highly plastic responsiveness to their environment. But arthropods such as dragonflies and praying mantids visually track prey as efficiently as cats do, whether they are conscious (as a few researchers claim[18]) or not.[19] So should we say not merely that conscious animals can have unconscious perceptions, but also that animals can perceive objects in their environment without having consciousness at all? Or was Aristotle (2017, 413b21–24) right to restrict perception to animals that experience desire, pleasure, and pain? The approach I am taking is consistent with both positions and I shall not attempt to decide between them here. But if sense perception can*not* occur without the capacity for desire, pleasure, and pain, then it is not merely its plasticity and its connection with goal-directed behaviour that account for the distinction between visual perception proper and the sensitivity to light of roundworms, or between the perception of heat and cold and the operation of a thermostat, it is these features of animal life as well.

Turning to the difference between conscious and unconscious sense perception, this does not simply consist in consciousness itself. It has further features, of which two play an especially important part in shaping the concept of perception.

First, conscious sense perception can be pleasant or painful, interesting or dull. This is not to say that only the objects we perceive consciously can make us feel excited, happy, or distressed. Unconscious perception can have these effects as well. It means that consciously seeing or hearing something can itself be hedonically or emotionally coloured, as when a cat enjoys the smell of catnip, or when a child is fascinated by an ugly face. Second, conscious sense perception is a source of factual

knowledge. For example, if you consciously perceive the green colour of an apple, and know that it is an apple you are seeing, you will normally see, and therefore know, *that it is green*. Hence, the fact that it is green can be your reason for choosing it or rejecting it, and in general for believing or wanting something, and for acting, or not acting, in a certain way. By contrast, the influence of unconscious perception on thought and behaviour normally bypasses factual knowledge. Phillip Merikle has studied this experimentally. The way he puts it is that conscious perception allows subjects to use information to guide their actions, so that they can follow instructions, whereas unconscious perception normally does not (Merikle and Joordens 1997a, 1997b).[20]

The case of blindsight also illustrates this point. Blindsight subjects have had part of their striate cortex destroyed, either by injury or because it was surgically removed to treat disease. As a result, they lose conscious visual experience in part of their visual field. But forced-choice experiments have shown that they remain able to identify the locations of point sources of light within the 'blind' part of their visual field, and even the shape of a stimulus, such as a circle or a cross. The subjects deny that they see the stimulus, and think of their answers to the forced-choice tasks as guesses, because as far as they can tell, their conjectures are made at random, or at least from uncertain indications, such as an impulse to plump for one answer rather than another. A guess does not need to be perfectly random or completely uninformed. Hence, it may not be an accident that a guess is right. But it is generally agreed that what is guessed or plumped for is not known.[21]

In sum, the distinction between conscious and unconscious perception is of critical importance in the theory of perception. If we are mindful of it, we can explore the ways in which perception, action, knowledge, and pleasure are related, freed from the erroneous ideas that perception is conscious by definition and that a perception is a sense-impression with a specific kind of cause.[22]

Notes

1 This provides a necessary condition for perceiving an 'external' object. As with the causal theory of intentional action, also advanced in the early 1960s, the task of supplementing this necessary condition to provide sufficient conditions was in due course widely acknowledged to be intractable.

2 Grice and Strawson (1956) defend the distinction against Quine's criticism. Glock (2010) defends it against Timothy Williamson's criticism. See also Glock (2017).

3 The fact that the modern theory is an exercise in conceptual analysis accounts for, but it does not justify, the narrow range of evidence they consider. The arguments against the causal theory I set out below draw on scientific and historical evidence.

4 I shall follow Grice here, and assume that the free use he makes of the terms 'impression' and 'sense-impression' is acceptable. Where Grice uses a sentence

of the form 'It seems to S as if he can perceive (see, feel, etc.) . . . ', Strawson uses 'It sensibly seems to X just as if he was seeing (feeling, etc.) . . . '; others use, 'X has a visual (tactile, etc.) experience as of . . . '; and so on. There are evidently different ways in which these kinds of sentences can be interpreted or understood. For example, one might interpret Strawson's formulation as referring to an experience that represents the subject as having an experience, but this is not what Strawson has in mind. If I caught sight of myself in a mirror with a startled expression, I could look to myself just as if I was seeing a ghost. But this is an unusual kind of case. 'It sensibly seems to X just as if he was seeing Y' is not meant to refer to an experience that represents X as seeing Y. Nor is it meant to prejudge theoretical questions about perceptual experience—for example, what phenomenal qualities or what kinds of content it can have.

5 A tip: if smelling X by smelling Y—e.g., smelling lamb chops by smelling the vapour they emit—involves a causal relationship between X and Y, it does not follow that it also involves a causal relationship between Y (or X) and the experience of smelling X.

6 Strictly speaking, these are explanations of the fact that he *could not* see the first pillar and the fact that he *could* see the second pillar—as opposed to *did not* and *did*. But we might explain the fact that someone *did not* see something in this way if it was relatively salient: There was a presumption that he would see it if it was visible, but it was not visible. And we might explain the fact that someone *did* see something in this way if there was a presumption that he would not see it, for example, because it was behind him. Explanations can be highly context-sensitive, in this way.

7 Strawson concurs. He refers in this connection to 'a specific concept, intrinsic to the naïve concept of perception from a-point-of-view, of the causal conditions under which a thing is accessible to perception, namely, that of being within unobstructed range of the relevant organ', and he argues that this should not be assimilated to 'the general idea of causal ways or means whereby a material object is causally responsible for producing the experience of perceiving it' (Strawson 1974, 90).

8 The same *petitio* invalidates other arguments modelled on Grice, such as in Pears (1976) and Lowe (2008).

9 In his response to Grice in the original Aristotelian Society symposium, White writes: 'Grice's argument [. . .] provides not the slightest evidence for the truth of what he admits is a claim which the causal theory of perception must necessarily make, namely that "perceiving a material object involves having or sensing a sense-datum"'. (White 1961, 156). White adds, 'I am not saying that the claim is false but only that Grice has given no evidence whatsoever for its truth'. By contrast, I *am* saying that the claim is false.

10 For a review of the evidence, see Merikle et al. (2001). See also Peirce and Jastrow (1884).

11 On sentence-nominals, see Hyman (2001); on perfect and imperfect nominals, see Vendler (1967).

12 The theory that an act is a movement of the agent's body caused in the right way by their mental states was defended in a similar way in the 1960s and 1970s. I criticise the assumption that an act is a movement of the agent's body in Hyman (2015, ch. 3).

13 Two difficulties about the interpretation of these passages have muddied the waters. Lowe (2008, 103–104) interprets the first passage as advancing an externalist view of the content of perceptual experience that is consistent with the causal theory and Child (1994, ch. 5) alleges that the second passage relies on an intuitive grasp of what makes two sorts of states of affairs

distinct, or what makes the nature of one case different from the nature of another, and that the claim in this passage is consistent with the causal theory of perception, if kinds of states of affairs are defined, and their natures are distinguished, in a certain way. But it is not difficult to identify the claim that Grice and Strawson are committed to, which Snowdon means to deny, namely the claim that a particular sense-impression may be either a perception or an hallucination depending on how it is caused. The remark that perception and hallucination are 'of a quite different nature' is to be understood as contradicting this claim. The way of spelling this out I prefer is to say that they are 'of a quite different nature' inasmuch as the former is a relation whereas the latter is not. For example, regardless of its content, *X*'s perception of an oasis is a relation between *X* and an oasis, whereas *X*'s hallucination is not. (By 'regardless of its content' I mean whether it looks to *X* as if there is an oasis in front of him or something else, say, a mirage.) I shall assume that this—as opposed to the externalist idea about content—is what Snowdon's reference in the first passage to a state of affairs that 'involves surrounding objects' and one that is 'intrinsically independent of surrounding objects' is intended to convey.

14 However, X and Y are distinct if X can exist or occur without Y existing or occurring, even if the reverse is not also true. For example, the conception and the birth of a child are distinct events.

15 Unsurprisingly, philosophers interested in defending the causal theory of perception against disjunctivist detractors have insisted that the state of affairs a looks-ascription describes or is made true by is the same, whether the subject is perceiving or hallucinating. See, for example, Lowe (2008, 109).

16 See Lindberg (1976, ch. 4).

17 I defend this conception of knowledge in detail in Hyman (2015, ch. 7).

18 See, for example, Barron and Klein (2016).

19 See Land (1992). On animal consciousness, see also Glock (2021).

20 Dretske (2006) defends the stronger claim that unconscious perception cannot make a fact about the object perceived available to guide thought or action.

21 Lowe (1996, 104) says that blindsight subjects are right to deny that they see the stimulus, on the grounds that they do not have a 'visual experience' of seeing it, while rejecting their claim to be guessing, because they are not merely accidentally right. In my view, this is exactly the wrong way round. If we reserve the term 'visual experience' for a *conscious* visual experience, then it is true that the blindsight subject does not have a visual experiences of the stimulus, but it does not follow that he does not see it. Alternatively, if seeing an object is, by definition, a visual experience, albeit in some cases an unconscious one, he does have a visual experience of the stimulus. But blindsight subjects are right to insist that they are guessing, for the reason stated in the main text.

22 I am grateful to Simon-Pierre Chevarie-Cossette, Hanjo Glock, Scott Sturgeon, Natalia Waights-Hickman, and the editors of this volume for helpful comments on previous drafts.

References

Aristotle. (2017). *De Anima*. Translated by C.D.C. Reeve. Indianapolis, IN: Hackett.

Armstrong, D.M. (1993). *A Materialist Theory of the Mind*, revised edition. London: Routledge.

Austin, J.L. (1962). *Sense and Sensibilia*. Edited by G.J. Warnock. Oxford: OUP.

Barron, A.B. and Klein, C. (2016). What Insects Can Tell Us about the Origins of Consciousness. *Proceedings of the National Academy of Sciences of the United States of America* 113: 4900–4908.

Chemero, A. (2009). *Radical Embodied Cognition*. Cambridge, MA: MIT Press.

Child, W. (1994). *Causality, Interpretation, and the Mind*. Oxford: OUP.

Dretske, F. (2006). Perception without Awareness. In Gendler, T.S. and Hawthorne, J. (eds.), *Perceptual Experience*. Oxford: OUP, pp. 147–180.

Feinberg, T. and Mallatt, J. (2017). *The Ancient Origins of Consciousness*. Cambridge, MA: MIT Press.

Glock, H.-J. (2010). From Armchair to Reality. *Ratio, New Series* 23: 339–348.

Glock, H.-J. (2017). Impure Conceptual Analysis. In Overgaard, S. and d'Oro, G. (eds.), *The Cambridge Companion to Philosophical Methodology*. Cambridge: CUP, pp. 83–107.

Glock, H.-J. (2021). Animal Consciousness—A Limit of Language? In Appelquist, H. (ed.), *Wittgenstein and the Limits of Language*. Abingdon: Routledge, pp. 197–222.

Goldman, A.I. (1976). Discrimination and Perceptual Knowledge. *The Journal of Philosophy* 73: 771–791.

Grice, H.P. (1961). The Causal Theory of Perception. *Proceeedings of the Aristotelian Society* 35(suppl.): 121–152.

Grice, H.P. and Strawson, P.F. (1956). In Defence of a Dogma. *Philosophical Review* 65: 141–158.

Hyman, J. (2001). -ings and -ers. *Ratio* 14: 298–317.

Hyman, J. (2015). *Action, Knowledge, and Will*. Oxford: OUP.

Land, M.F. (1992). Visual Tracking and Pursuit: Humans and Arthropods Compared. *Journal of Insect Physiology* 38: 939–951.

Lindberg, D. (1976). *Theories of Vision from al-Kindi to Kepler*. Chicago, IL: University of Chicago Press.

Locke, J. (1997). *An Essay Concerning Human Understanding*. Edited by R. Woolhouse. London: Penguin.

Lowe, E.J. (1996). *Subjects of Experience*. Cambridge: CUP.

Lowe, E.J. (2008). Against Disjunctivism. In Haddock, A. and Macpherson, F. (eds.), *Disjunctivism: Perception, Action, Knowledge*. Oxford: OUP, pp. 95–110.

Martin, M.G.F. (1997). The Reality of Appearances. In Sainsbury, M. (ed.), *Thought and Ontology*. Milan: Franco Angeli, pp. 77–96.

Merikle, P.M. and Joordens, S. (1997a). Parallels between Perception without Attention and Perception without Awareness. *Consciousness and Cognition* 6: 219–236.

Merikle, P.M. and Joordens, S. (1997b). Measuring Unconscious Influences. In Cohen, J.D. and Schooler, J.W. (eds.), *Scientific Approaches to Consciousness*. Mahwah, NJ: Erlbaum, pp. 109–123.

Merikle, P.M., Smilek, D. and Eastwood, J.D. (2001). Perception without Awareness: Perspectives from Cognitive Psychology. *Cognition* 79: 115–134.

Pears, D.F. (1976). The Causal Conditions of Perception. *Synthèse* 33: 25–40.

Peirce, C.S. and Jastrow, J. (1884). On Small Differences in Sensation. *Memoirs of the National Academy of Sciences* 3: 73–83.

Ryle, G. (2009). *The Concept of Mind*, 60th anniversary edition. London: Routledge.

Snowdon, P. (1980). Perception, Vision, and Causation. *Proceedings of the Aristotelian Society* 81: 175–192.

Snowdon, P. (1990). The Objects of Perceptual Experience. *Proceedings of the Aristotelian Society* 64: 121–166.

Snowdon, P. (1998). Strawson on the Concept of Perception. In Hahn, L.E. (ed.), *The Philosophy of P. F. Strawson*. La Salle, IL: Open Court, pp. 293–310.

Snowdon, P. (2005). The Formulation of Disjunctivism: A Response to Fish. *Proceedings of the Aristotelian Society* 105: 129–141.

Strawson, P.F. (1959). *Individuals*. London: Methuen.

Strawson, P.F. (1974). Causation in Perception. In *Freedom and Resentment*. London: Methuen, pp. 66–84.

Strawson, P.F. (1979). Perception and Its Objects. In Macdonald, G.F. (ed.), *Perception and Identity*. London: Macmillan, pp. 41–60.

Strawson, P.F. (1992). *Analysis and Metaphysics*. Oxford: OUP.

Strawson, P.F. (1998). Reply to Paul Snowdon. In Hahn, L.E. (ed.), *The Philosophy of P.F. Strawson*. Chicago, IL: Open Court, pp. 311–314.

Strawson, P.F. (2000). Entity and Identity. In *Entity and Identity and Other Essays*. Oxford: OUP, pp. 21–51.

Sturgeon, S. (2008). Disjunctivism about Visual Experience. In Haddock, A. and Macpherson, F. (eds.), *Disjunctivism: Perception, Action, Knowledge*. Oxford: OUP, pp. 112–119.

Vendler, Z. (1967). Facts and Events. In *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.

White, A.R. (1961). The Causal Theory of Perception. *Proceedings of the Aristotelian Society* 35(suppl.): 153–168.

Wittgenstein, L. (1958). *Philosophical Investigations* [=PI]. 2nd edition. Translated by G.E.M. Anscombe. Oxford: Basil Blackwell.

Part III

Animal minds

9 Understanding animal minds

Between hermeneutics and hydraulics

Markus Wild

With a Little Help from My Friends

—The Beatles

Introduction: two idioms

In his study *The Psychoanalytic Movement* (1985), Ernest Gellner points out that Sigmund Freud used or rather fused two very different idioms in order to describe the human psyche, namely a hermeneutic or interpretive vocabulary and a quasi-biological or hydraulic one. Accordingly, interpretations of Freud fall into two rival camps, which might be called 'the Materialist and the Hermeneutic':

> The former will treat the 'scientific' aspiration to link psychoanalytic theory to biology and to a theory of drive and to the physiological bases of our mental life. The latter, by contrast, will concentrate on the closeness of psychoanalytic theory and practice to life as actually lived. It is important for them that the theory remains close to the realm of *meanings* experienced by individuals.
>
> (Gellner 1993, 110)

When interpreting and understanding animal minds, the difference between Hanjo Glock's approach and mine somewhat resembles the rivalry as to which interpretation of Freud's theory of the unconscious is accurate. One only would have to replace 'psychoanalytic theory' in the quotation above by 'animal behaviour theory' and read 'life' and 'meanings' in close proximity to linguistic practices. While Glock accepts the lead of our linguistic practices and while he pursues a capacity approach favouring what he calls a 'hermeneutic ethology', I am rather sceptical of our linguistic practices and I pursue a representationalist approach favouring what I will call, for the purpose of this contribution, a 'hydraulic ethology'. However, despite our differences, we both think that non-human and non-linguistic animals think, act, and reason. Our differences, therefore, are—as Rorty once put it to Habermas—'merely philosophical'. In the first parts of my contribution, I will focus on some

DOI: 10.4324/9781003202929-13

of our differences; in the last part, I will show how our agreement on animal minds teams up against a common enemy. In a nutshell (and in allusion to the famous Kantian dictum), my main claim is that hermeneutic ethology without hydraulic ethology is empty.

The hermeneutic and the hydraulic ethologists

Glock claims that, in order to counter the threat of content indeterminacy in non-linguistic subjects, the hermeneutic ethologist will be able to determine content by exclusion, i.e. by ruling out alternative attributions by scientific experiment and by philosophical argument (Glock 2020, 102). More generally, the hermeneutic ethologist holds that the justified attribution of mental states to animals starts by attributing actualisations of mental capacities in reasonable ordinary language terms (such as 'The dog sees that the bone is in the bowl'), which is followed by ruling out alternative interpretations of the animal behaviour.

By contrast, the hydraulic ethologist thinks that a mental capacity is always constituted by the function of some biological mechanism guiding complex and intelligent animal behaviour. More specifically, every genuine mental capacity is a capacity for producing mental representations. To quote Fred Dretske: 'All mental facts are representational facts, and all representational facts are facts about informational functions' (Dretske 1995, xiii). According to this view, there must be some mental machinery in place (some pushing, rotating, and puffing; some heaving, displacing, and steaming; some computing, converting, and calibrating) which constitutes the actualisation of a mental capacity, such as the actualisation of the capacity for visual perception that fits the description, 'The dog sees that the bone is in the bowl'. The hydraulic ethologist is happy to admit that the dog sees a bone and a bowl, but she thinks that the contents of what the dog sees are not constituted by the properties of the bone and the bowl alone, but rather by the biological machinery that allows the dog to visually represent a bone in a bowl.

To this line of thought the champion of the capacity approach will probably reply that the hydraulic ethologist makes some kind of category mistake. The mind, he says, is not 'a special kind of thing, whether it be a mental substance, as in dualism, or the brain, as in materialist monism' (Glock 2019b, 147). Rather, the mind is a potentiality or power, and to have a mind 'is to have a range of cognitive, volitional, and affective powers' (Glock 2019b, 149). Well, the hydraulic ethologist I am representing is not committed to a Cartesian picture of the mind at all, neither to Cartesian dualism nor to Cartesian materialism.[1] The hydraulic ethologist is a Cartesian machinist. According to Descartes, non-human animals are quite literally divine machines, i.e. machines created by the omnipotent creator of all things, and these machines are essentially collections of special-purpose mechanisms for self-preservation.

From the point of view of the hydraulic ethologist, Descartes was on the right track, except for three important caveats: (i) non-human animals are not divine machines created by God, but biological machines created by Darwinian evolution; (ii) non-human animals are collections not only of special-purpose mechanisms, but also of general-purpose mechanisms (mostly for self-preservation); and (iii) human animals are no exception to (i) and (ii). While (i) is clearly a post-Darwinian point, points (ii) and (iii) were already made by Descartes' greatest heretical follower, Spinoza.[2] Thus, the interesting point about Descartes in this context is not substance dualism or properties of the mental, but rather the revolutionary idea that living beings are composed of mechanisms with purposes or functions.

Part of the animal machine's special-purpose and general-purpose mechanisms are mechanisms for producing mental representations of the animal's physical and/or social environment and of the animal's body and/or mind. These mental representations have accuracy or truth conditions, because they have (remember Dretske's quote) informational functions. They have the function of carrying certain kinds of information about the animal's physical and/or social environment and about the animal's body and/or mind. If the mechanisms work according to their functions and if the mental representations are produced in suitable conditions, the upshot will be accurate or true representations; if the mechanisms malfunction or are forced to work under non-suitable conditions, the resulting mental representations will be inaccurate or false, they will carry misinformation and, thus, misinform the animal. Now, since animals are (to squeeze a complicated story into a simple headline) biological machines created by Darwinian evolution, both evolutionary history and life history are essential for having a mind.

That having a past is essential for having a mind makes sense intuitively if we focus on the capacities to store information in memory, to learn from memory, and to remember general or specific facts, events, and individuals. The third great early modern rationalist, Leibniz, claims that only those things should be said to have a soul or a mind 'whose perception is more distinct and accompanied by memory' (Leibniz 1991, 19), for memory provides 'a kind of connectedness to souls which resembles reason' (Leibniz 1991, 20). The capacities of souls or minds in this sense are not limited to non-human animals, but form a great part of the mental life of human animals: 'Men function like beasts insofar as the connections among their perceptions come about only through the agency of memory' (Leibniz 1991, 20). In sum, animal minds have perceptions that are distinct enough to give rise to memory and sensation (in addition, animals can feel pleasure and pain). If Leibniz is right and memory is a constitutive capacity for having a mind, it follows that having a past is constitutive for having a mind.

In the teleosemantic picture (Dretkse 1995; Millikan 2004; Neander 2017), I am drawing on behalf of the hydraulic ethologist, animal minds are literally made of biologically evolved mechanisms for producing mental representations. Mental representations have the functions of carrying certain kinds of information, and having an evolutionary and individual past is essential for having a mind (this is reflected in the importance of memory for minds). Of course, this picture has a well-known consequence that many philosophers consider to be intolerable, namely that a spontaneously generated creature that behaves and looks exactly like me or my dog would not be able to exercise any mental capacities whatsoever, since it lacks a past. Let us adjust a well-known thought experiment for our present purpose (Davidson 1987, 443): Lo, lightning strikes a dead tree in Swabia and Hanjo is running nearby! His body is reduced to its elements, while entirely by coincidence (and out of different molecules) the tree is turned into a physical replica of Hanjo, the Swamp-Swabian. It moves exactly as Hanjo did; it seems to recognise his friends, and it appears to return their greetings in German. It moves into his house and seems to write articles on animal minds. No one can tell the difference. However, there is a difference: Hanjo's replica cannot recognise his friends; it cannot recognise anything, since it never cognised anything in the first place. It cannot know his friend's names, it cannot remember his house. Indeed, the Swamp-Swabian cannot be said to mean anything by the sounds it makes, nor to have any thoughts.

Again, the thought experiment is the most intuitive with regard to memory. The Swamp-Swabian cannot remember having taught at the University of Reading or written the Wittgenstein Lexicon. It simply did not exist at that time. It only seems to remember these things, yet seeming to remember is not to be confused with remembering. The metaphysical point of the thought experiment is the claim

> that people who are in all relevant physical respects similar (or identical in the necktie sense) can differ in what they mean or think, just as they can differ in being grandfathers or being sunburned. But of course there is *something* different about them, even in the physical world; their causal histories are different.
>
> (Davidson 1987, 452)

Pace Davidson, the same goes for animals. Animals who are in all relevant physical respects similar (my dog Titus and Swampdog) can differ with regard to what they perceive, feel, remember, or think. Therefore, in the teleosemantic picture I am drawing here, the mind is not a material thing *tout court* (e.g., the brain), but rather it is an assemblage of mechanisms which have a past and the function of which is to produce mental representations.

The teleosemantic capacity approach

I have to admit that I am not sure whether the teleosemantic picture of the mind counts as an instance of what Glock calls the capacity approach, according to which the mind is a potentiality or power. Probably, the hermeneutic ethologist will protest and reply that neither the causal machinery at work nor the environmental conditions under which a power is actualised, nor the biological make-up of the organism, nor the physiological vehicles sustaining the cognitive power will determine whether an animal possesses mental powers in the first place. Well, the hydraulic ethologist does not claim that a power to produce mental representation with informational functions is identical to the actual exercise of that power or the environmental conditions sustaining it, or the representational vehicles. Mechanisms for producing mental representations may not be functional in the first place. A dog born blind has the potential capacity to see his environment (that is why we call him 'blind'), but he is not able to exercise this specific mental capacity due to a congenital defect. In other words, the dog has the relevant machinery for seeing, but the machinery is *not functioning*. A dog blind in one eye will be able to see objects in his environment; however, he will experience severe setbacks in locating these objects in egocentric space. In other words, the dog has the relevant machinery for seeing, but the machinery is partly *malfunctioning*. What is crucial in the teleosemantic picture, however, is the biological make-up of the organism—if by 'biological', one means 'functionally designed by the evolutionary and individual history of the animal'. Again, two animals who are in all relevant physical or physiological respects similar can differ in what they perceive, feel, remember, or believe. What is more, they could even be members of two completely different species if we were to find them on two different planets with two completely different evolutionary origins, for example. Therefore, the animal mind as defined by the teleosemantic capacity approach is not a thing or stuff; it is the functionally defined capacity to produce mental representations.

I think that a great advantage of the teleosemantic capacity approach to animal minds that the hydraulic ethologist is promoting consists in the circumstance that it cannot be targeted by the *virtus-dormitiva* objection that Nietzsche raised against Kant's philosophical project:

> How are synthetic judgments a priori possible? Kant asked himself, —and what really was his answer? By virtue of a faculty, which is to say: enabled by an ability [Vermöge eines Vermögens] [. . .] But is that really—an answer? An explanation? Or instead just a repetition of the question? So how does opium cause sleep? "By virtue of a faculty," namely the virtus dormitiva—replies the doctor in Molière,
>
> *quia est in eo virtus dormitiva,*

> *cujus est natura sensus assoupire.*
> But answers like this belong in comedy …
> (Nietzsche 2002, 12–13; BGE 1, 11)

The *pure* capacity approach explains the capacities of animal minds by mental powers, and it explains the *actual* exercise of mental powers by virtue of potentialities. So, how does the dog see the bone in the bowl? Well, by virtue of his power to see objects, by a *virtus perceptiva*. This kind of explanation is empty or circular. Is it really true that appealing to capacities and potentialities constitutes good philosophy, even though it often leads to bad science, as Glock holds? I do not believe that empty or circular explanations lead to good philosophy, and I also do not believe that what leads to bad science leads to good philosophy. By contrast, the teleosemantic capacity approach is not vulnerable to the *virtus-dormitiva* type of objection, for it explains the mental capacities of animals by virtue of the informational functions of the evolved or learned mechanisms that produce mental representations. Of course, one could still object that mental representations are useless theoretical entities, but this is another issue. The important point for now is the following: While the pure capacity approach of the hermeneutic ethologist falls prey to the *virtus-dormitiva* objection, the teleosemantic capacity approach does not. That is one point for the hydraulic ethologist. In the next section, she will score some more by showing that the hermeneutic ethologist cannot do without hydraulic ethology.

Seeing objects: dogs, bones, bowls, and tables

In Glock's arsenal of examples, dogs and bones have a clear priority (e.g. Glock 2009, 43–44; 2010, 21f.; 2013, 228ff.; 2019a, 666ff.; 2019b, 111–112). His argument for animal mentality usually starts with the claim that higher animals with sense organs (such as dogs) are capable of perceiving objects and facts. Since perceiving that *p* implies either knowing that *p* or believing that *p*, his argument concludes that higher animals can know or believe that *p*. The most important premise in this line of reasoning is that: The reaction of higher animals to their environment can only be explained by a capacity to perceive that *p* (Glock 2009, 224). Thus, the hermeneutic ethologist argues that some complex behaviour of dogs and other higher animals can only be explained by the capacity for fact perception, and he does so by ruling out alternative interpretations. How is this supposed to work?

> Consider a dog who has been taught not to grab anything when it is lying on the table, but only when it is lying in its bowl. This dog now sees a bone on the table, but refrains from grabbing it and instead looks on, panting. Yet, as soon as the bone is placed in the bowl, the

> dog goes for it. This mundane sequence of events is not explained by the dog simply perceiving discrete objects – the bone, the table, and the bowl. It can only be explained in terms of the following opposition: (1) The dog sees at time t_1 that the bone is on the table; and (2) The dog sees at time t_2 that the bone is in the bowl.
>
> (Glock 2019b, 151)

Why is that so? Well, according to the hermeneutic ethologist, the dog can see all three objects at times t_1 and t_2. Since all the objects are co-present in the dog's visual field (as Glock seems to construe the situation), the perception of the conglomeration formed by bone, table, and bowl cannot explain the difference in its behaviour at t_1 and t_2.

This is not very convincing as it stands, for the dog might build new perceptual objects such as 'table-bone' (or 'tabley-bone') and 'bowl-bone' ('bowly-bone'), depending on spatial relations between bone, table, and bowl, or depending on their location in his visual field, or depending on partial occlusion.[3] The dog goes for 'bowl-bone', but not for 'table-bone', and as he approaches the bowl, his sense of smell and touch will do the rest. To be clear, my objection does not depend on the claim that the dog perceives, at t_1, that the bone has the property of being on the table and, at t_2, that the bone has the property of being in the bowl, because in this case we are back with perceiving that *p* (Glock 2010, 22). The objection is that the dog might just form new objects as a result of his individual learning history (while the general capacity to perceive objects is a result of the evolution of his sensory apparatus).

This 'new object objection' seems at first sight rather surprising, however there is some plausibility to it. First, my dog Titus treats cats very differently depending on whether they are inside or outside our house. A 'cat-in-da-house'[4] is a very different cat from a 'cat-outside-the-house'. Titus never chases a 'cat-in-da-house' (never mind the difference between my own cats and intruding cats), but he almost always chases 'cats-outside-the-house'. For an animal, the criteria of identity might be simply practical. Thus, recognizing an object that is over there in such and such a direction simply equals knowing how to respond to or use that object, given a context of practical concern. If two objects function in the same way when used the same way, they are the same; if they function differently or must be used differently, then they are different. So, for Titus all cats are roughly the same thing, with the exception of the 'cats-in-da-house'. A 'cat-in-da-house' is of a different practical kind from other cats.[5] On this view, a bone on the table is a very different thing from a bone in the bowl to any well-trained dog. While the 'table-bone' is something to be shunned, the 'bowl-bone' is something to be grabbed.

Second, the situation of the dog perceiving the bone on the table and (then) in the bowl might be more like seeing a happy face and an angry

face than seeing that this face is happy and that the second face is angry. Dogs are good at recognizing emotions in human faces (Correia-Caeiro et al. 2020). But this does not mean that they are good at identifying emotions as such. A dog is certainly able to see a happy face and react accordingly (panting and wagging), but this does not imply that the dog is able to see that the face (or the person) is in a state of happiness. Happiness as such is not something the dog is able to detach in perception from faces. Nevertheless, the dog is capable of distinguishing happy faces from angry ones. This indicates that dogs perceive complex or compound objects rather than structured facts.[6] Given these insights from comparative psychology, a dog is able to see Markus's happy face or Hanjo's happy face, but this does not imply that the dog can think both that Markus is happy and that Hanjo is happy, since the dog does not need the underlying ability to think about being happy. For the dog, seeing happy faces or angry faces does not amount to seeing facts but rather to seeing complex objects (happy-faced-master at t_1 and angry-faced-master at t_2). For the dog, 'happy-faced-master' and 'angry-faced-master' on the one hand and 'tabley-bone' and 'bowly-bone' on the other hand might be fully determinate concrete objects. If the dog is able to distinguish between complex objects such as 'tabley-bone' at t_1 and 'bowly-bone' at t_2, then it is not the case that the dog's behaviour can only be explained in terms of the opposition that the hermeneutic ethologist put forward above. For us to decide whether the hermeneutic ethologist's interpretation of the dog's behaviour or the new object account is correct, we need to know more about the dog's visual field, his way of individuating objects across different sense modalities, his ability to follow moving objects, his ability to distinguish objects, etc. Since both the hermeneutic ethologist's and the new object interpretations can account for the dog's behaviour, the question cannot be answered as long as we do not take the dog's mental machinery into consideration. In other words, we need help from the hydraulic ethologist. As I will show in the following, Glock himself relies on her help in his master argument for animal thought.

Returning to the hermeneutic ethologist's interpretation of the dog who has been taught not to grab anything on the table but only when it is lying in the bowl, the first objection to his fact-perception account he considers is the stimulus-response-interpretation of the behaviourist:

> But what sort of stimulus? Is it purely proximal and physiological, like the pain stimulus to which even an oyster will react? This behaviourist fairy tale ignores the distinction between lower animals and higher ones like dogs, dolphins, or primates, which possess a range of different sense organs and corresponding sensory centres in the brain. Primates, at least, score well in the standard tests for object permanence and object identification [. . .] They not only perceive

> the same distal object *x* in spite of altering proximal stimuli, they can also keep track of *x*, even as *x* moves in space and changes some of its properties. [. . .] The alternative to the behaviourist tale is to admit that the dog's reaction is not just to a proximal stimulus, but to information about distal objects acquired through vision.
>
> (Glock 2019b, 152; cf. 2013, 228ff.)

The distinction between higher and lower animals is drawn by reference to 'sense organs and corresponding sensory centres in the brain' and by reference to the distinction between 'a proximal stimulus' and 'information about distal objects'. In other words, in order to refute the behaviourist tale, the hermeneutic ethologist relies on the internal mental machinery of some animals that allow them to systematically ignore changes in the proximal stimulus in order to keep track of a distal object. But this is just what the hydraulic ethologist claims: There must be some mental machinery (some pushing, rotating, and puffing; some heaving, displacing, and steaming; some computing, converting, and calibrating) that constitutes the actualisation of a mental capacity for visual perception.

In order to keep track of a distal object, the higher animal needs some mental machinery that allows for objectification. According to one fellow hydraulic ethologist, Tyler Burge, objectification 'is a process that systematically contrasts phenomena that encode proximal stimulation, at various levels of abstraction, and phenomena that represent specific environ-mental entities' (Burge 2014, 400; cf. Burge 2011). Objectification, therefore, is the distinctive feature that separates genuine representational capacities like perception from the mere sensitivity of simpler animals. More specifically, what marks an animal's mental state as perceptual is some mental machinery that contrasts registrations of proximal stimulation with states that specify elements of the environment beyond the sensory receptors. It is by means of perceptual constancies that a higher animal keeps track of distal properties of its environment, which allows it to mentally represent distal objects (Burge 2010; Schulte 2021).

Animals do have the necessary perceptual constancy mechanism that allows them to track shape, size, luminance, etc. (Feng et al. 2017).[7] If the dog's master takes the bone off the table and puts it into the bowl on the floor, we have good reasons to think that the dog will be able to track the object we call 'bone'. Nevertheless, he may perceive, at a certain point in time, the compound object I named 'bowl-bone', as this object is of practical significance for the dog's behaviour in that it positively affords grabbing the bone in the presence of his master. Also, we may still admit that the dog approaches the 'bowl-bone', and as he approaches the bowl, his sense of smell and touch will do the rest, we do not need objective perception *sensu* Burge to explain that the dog grabs

the bone (rather than the 'bowl-bone'). In other terms, that the dog grabs the bone does not warrant the conclusion that the dog actually sees the bone (rather than the 'bowl-bone'), nor, for that matter, that the dog sees that *p*. What provides us with some warrant here is the hydraulic ethologist's idea that the dog visually represents the bone and that he does so by perceptual constancies, enabling him to keep track of distal objects. In sum, to get off the hook of the new object objection and in order to avoid the false spell of the behaviourist tale, the hermeneutic ethologist needs a little help from his hydraulic friend. The point is very simple. Only the hydraulic ethologist will be able to show that the dog's visual system does indeed track ordinary objects (e.g. by Burge's constancy mechanism allowing for distal object representation and, thus, for objectification) and not complex new objects (such as 'tabley-bone' or 'bowl-bone'). Moreover, such representational mental machinery speaks against the behaviourist tale. Hermeneutic ethology by itself will not be able to make sense of the distinction between animals that have sense organs and corresponding sensory centres in the brain that allow them to access information about distal objects through the blooming, buzzing confusion of proximal stimulation and other living creatures. Hence, Glock's master argument for fact perception and thought in animals cannot take off without a little help from his hydraulic friend.[8]

The problem of colour

A crucial feature of object perception that the capacity approach is missing is colour vision. I submit that any creature that is able to visually perceive distal objects sees them in some achromatic and/or chromatic colours. Objects do have surfaces and surfaces must necessarily be (chromatically or achromatically) coloured. Therefore, if an animal has the capacity to produce mental representations of objects (in contrast to mere registration of proximal stimulation), this animal must be able to perceive achromatic and/or chromatic colours.

Consider birds and mammals. Birds started out as diurnal animals, mammals as nocturnal animals. Colour vision is an asset for diurnal animals and a nuisance for nocturnal ones. That is why many birds have good colour vision and mammals rather poor colour vision. While most bird species have three or four cone-cell types, most mammals have fewer. Primates (including humans) are the exception as they do have very good colour vision, but since birds and primates have a very different evolutionary past and since colour vision serves different purposes for them, we have to assume that there are important differences between the colour perception of, say, dogs, humans, and chickens. Phenomenologically speaking, we only have access to one kind of colour experience, namely our own way of perceiving colours. Given the difference in the number of cone-cell types (two in dogs, three in humans, four in chickens), the

different evolutionary purposes of colour vision, dogs and chickens can probably see objects in colours we cannot even imagine. Both the relatively limited colour space of the dog and the relatively extended colour space of the chicken will contain hues beyond human experience. Almost nothing the hermeneutic ethologist can offer will be helpful for understanding the way animals perceptually represent colours.

Colour discrimination tasks will not help: They can only tell us that animals are able to discriminate between surfaces we, say, call 'red' and 'yellow', but they do not answer the question how animals see the colours we call 'red' and 'yellow'. Chickens, which are tetrachromats, are equipped with four different types of cone cells and three opponent processes, while humans have only three different types of cone cells and two opponent processes. Chicken colours are not merely more fine-grained, nor do they just see beyond the spectrum that is visible to us, but since they have three opponent processes, we also have to conceive of chicken colours as being removed from the colour experience that is at our disposal with our red-green, blue-yellow hue palette. Thus, since they are able to mix up different hues, the space of chicken colour is not only broader, but also really different from ours (Thompson 1992; Matthen 1999, 2018). So, all that colour discrimination tasks can tell us is that an animal is capable of distinguishing colours, and nothing more. Only the mechanisms underlying the representation of coloured surfaces (such as opponent processing) allow us to appreciate that the ways objects look to chickens, dogs, and humans are indeed substantially different from each other.

At this point, Glock might object that the example of colour is misleading given his explanatory purposes. His master argument is all about the perception of objects and the perception of facts. What the table, the bone, or the bowl look like to the dog or the chicken is entirely beside the point, because what matters is whether the dog sees that the bone is in the bowl independently of the colour of these objects. Fair enough! However, my point is that the capacity approach does not allow us to say anything about the ways animals perceive colours without backup from the hydraulic ethologist. Only the machinery of colour vision (such as cones and opponent processing) will help us to understand the enormous variety and disunity of colour vision in the animal world. Since colour is a necessary ingredient of object perception, the hermeneutic ethologist leaves out something important about object perception. Again, the hermeneutic ethologist cannot make sense of object perception in animals without a little help from his hydraulic friend.

The additive and the transformative interpretation of rationality

In the last section, I will show how the hermeneutic and the hydraulic ethologist can join forces to stand up against a common enemy. The

enemy is the so-called 'transformative' interpretation of the rational animal. So, what is the 'transformative' interpretation?

According to Aristotle, human beings are rational animals, while non-human animals are non-rational creatures capable only of vegetative processes, sensation, and locomotion. According to the 'additive' interpretation, humans possess the vegetative, sensitive, and locomotive capacities of animals and, in addition, they possess rationality. The opposing 'transformative' interpretation claims that rationality is not an add-on to the animal powers; rather, rationality transforms the mental capacity we share with non-rational animals. Humans realise the animal capacities in a rational way that fundamentally transforms the vegetative, sensitive, and locomotive capacities, while animals realise the same capacities in an entirely different, non-rational way. Glock and I both oppose this picture of the relation between human and non-human animal minds.

Glock has characterised the transformative approach in the following manner:

> In short, 'objective experience, i.e. genuine perception', depends on concept-possession and, hence, on rational/linguistic faculties. Because of their (actual or presumed) lack of concepts, animals cannot perceive the world in a sense of 'perception' that applies univocally to humans as well.
>
> (Glock 2019b, 156)

Glock holds that the transformative interpretation does not stand scrutiny, (i) because it does not follow from the claim that animals lack concepts that they cannot perceive what creatures with concepts can perceive; (ii) because there is no sense of the word 'perceiving' that does not apply unequivocally to humans and animals; and (iii) because the transformative interpretation is committed to an insane form of holism (Glock 2019b, 156–157). Alas, I think that Glock misconstrues the claim of the transformativists and that his objections misfire. Take McDowell's word for it. In contrast to Robert Brandom, McDowell does not have any problem with attributing to animals a capacity to be 'on to things' in their environment:

> I see no reason to think the knowledge of animals without conceptual capacities is only loosely so called, just because it is not the interesting kind of knowledge that the Sellarsian conception fits. I see no reason to think there is nothing really there but actualisations of responsive dispositions. The implication [*of Brandom's position*] is that, say, a cat's awareness of the prey it stalks is no more genuinely a case of awareness than is an 'awareness' of the presence of moisture shown by iron filings in rusting. This is the kind of thing

> nobody but a philosopher would suppose. (Descartes, perhaps.) I want no truck with it.
>
> (McDowell 2002, 104)

Even though Glock's reconstruction and criticism of the transformative interpretation are on the wrong track, I think that he has a very good reply to the transformativists.

Glock's view is that the rational and the non-rational realisations of the perceptual capacities depend on a distinction between the conceptual and the non-conceptual, as introduced by McDowell (1994). However, according to the best proponent of the transformative position, Matthew Boyle, this is not the real issue, for McDowell's claim that rational capacities are already at work in perception 'does not appear to depend on specific commitments about the nature of conceptual contents' (Boyle 2016, 535). While 'conceptual content' figures as a term for content, the attribution of which implies the engagement of rational capacities (the capacities to reflect on our reasons for belief and form a considered judgment), 'non-conceptual content' is a stand-in for content, the attribution of which implies the engagement of non-rational capacities (the capacities to react to environmental opportunities for survival). A dog, for example, can decide to grab the bone in the bowl; a human perceiver, in contrast, is able to pause and reflect on whether he should eat the last piece of chocolate on the table or whether something that looks and smells like chocolate really is chocolate.

McDowell's ambition

> is to raise a difficulty for any account that treats the cognitive powers of a rational perceiver in an additive way: as consisting of a not essentially rational power to perceive, whose acts of perception are inputs to a further and independent power to make reflective judgments.
>
> (Boyle 2016, 535)

The difficulty in question is, according to Boyle, that it raises 'interaction problems'. If rational powers are not involved in the deliverances of our sense organs, it becomes a mystery as to how a rational agent's perceptions can provide her with reasons to judge. In the same vein, if rational powers are not involved in the formation of our desires, it will be mysterious how a rational agent's desires can provide her with reasons to act. Thus, the transformativist does not hold a view according to which humans and animals cannot perceive the same things in the environment, i.e. that the cat is not aware of its prey over there. Rather, they hold that perceptions give us reasons for judgements and desires give us reasons for actions, while animals' perceptions and desires are only opportunities and imperatives for behaviour. While animals respond to

the biological imperatives of their bodies and environments, human lives 'come to embrace not just coping with problems and exploiting opportunities, constituted as such by immediate biological imperatives, but exercising spontaneity' (McDowell 1994, 115). Therefore, holism is not a real issue in this debate. (It is an issue in Brandom's inferential semantics, cf. Fodor and Lepore 2001.) The dispute is rather about whether and how the perceptions and motives we share with animals could provide us with reasons for judgements and actions as well as material for reflection. The transformativist holds that perceptions and motives must take the *form* of reasons from the outset for them to play this role at all. In contrast, the additive interpretation of Aristotle claims 'that it is possible to explain what is involved in our possessing the generically animal capacities for perception and motivationally efficacious desire without appealing to the specifically rational capacity for reflective judgment' (Boyle 2016, 537).

The transformativist claims that any account of perception in rational agents is bound to two fundamental commitments, which exclude the possibility of additive interpretations. The commitments are quite simple and fundamental: First, any explanation of a rational subject's *S* judging a judgment *J* must appeal to the reasons available to *S*'s reflective scrutiny. Second, that *S* has a perceptual experience of some object *o* being *F* can normally explain *S*'s judging *o* to be *F*. For example, when, after close inspection of the kitchen table, I judge that there is some chocolate left on the table, any explanation of my perceptual judgment must appeal to the reasons available to my reflective scrutiny. Moreover, my perceptual experience of chocolate on the table normally explains my judgement that there is chocolate on the table. However, according to the additive interpretation, a rational subject's perception does not present her with reasons available to her reflective scrutiny, because the perceptual capacities we share with animals stand outside the realm of rationality. Therefore, the defender of the additive interpretation faces an interaction problem, because he cannot have it both ways: He cannot stick to the two fundamental commitments mentioned above while maintaining that the non-rational interpretation of our perceptual capacities applies as well. As McDowell puts it, we are, according to the additive interpretation, 'peculiarly bifurcated with a foothold in the animal kingdom and a mysterious separate involvement in an extra-natural world of rational connections' (McDowell 1994, 78). For Boyle, this raises serious interaction problems that only the transformative interpretation can deal with.

Is Glock's capacity approach also susceptible to the interaction problem that motivates the transformative interpretation of Aristotle? I think that Glock has all the answers to the transformativist challenge. In my view, the hermeneutic ethologist and the hydraulic ethologist should both reject the conceptual oppositions that provide the background for the transformative interpretation. These oppositions include the opposition

between the normative space of reasons and the realm of natural laws, between rational animals and non-rational animals, and between an additive and a transformative interpretation of rational animals. The reason is that there exists, between the space of reasons inhabited by human beings and the realm of natural laws reigning over the vast empire of physics, the animal kingdom (Macdonald 2006). The teleosemanticist capacity approach holds that the functional explanations relevant for animal capacities to be found in biology are neither dependent on laws nor free of norms. Given the distinctiveness of functional explanations in biology and their normative character, there is at least a third realm, populated by plants, lower animals, and higher animals. According to the teleosemantic picture, animal minds are literally made of biologically evolved mechanisms for producing mental representations, and mental representations have the functions of carrying certain kinds of information. Thus, a higher animal's perceptual machinery has the function of representing objects in its environment (by means of mechanisms such as perceptual constancies and opponent processing). These representations provide the animal with reasons for having beliefs, memories, plans, and knowledge about its environment; moreover, they can provide an animal with reasons for action. There is no strict dualism between the normative space of reasons and the realm of natural laws.

It follows that we have no good reason to distinguish strictly between humans and animals, rational and non-rational creatures. The transformative interpretation presupposes this kind of anthropological difference without justification. However, if we can show that animals can perceive facts or believe that *p*, that animals can act upon reasons, then we have denied precisely this anthropological difference. Glock provides us with many convincing reasons for such a position (Glock 2009; 2010; 2019a).

Finally, the claims that animals can believe that *p* or that they can act upon and reflect on reasons dispel Boyle's worries about interaction problems: Since these perceptions are reasons for believing the perceptual content, our animal capacities are not beyond the rational capacity for making reflective judgements. On the contrary, they are ready in their form, so to speak, for reflection. In us humans, our ability to reflect on our own perceptions and desires in context merely adds to our animal capacities in that it takes certain desires and perceptions as input for judgements and actions, but it transforms other perceptions and desires by placing them in a larger context, structuring them in time, or re-evaluating them through prudence and wisdom. There is no good reason for distinguishing strictly between an additive and a transformative conception. The problem with the transformative approach of McDowell, Boyle, and others is that they present us with a false dilemma. The hermeneutic and the hydraulic ethologist should think jointly about the minds of animals independently of such false dichotomies.

Conclusion

I started my contribution by picking up Gellner's contrast between two very different vocabularies to describe the human psyche in Freud's theory, namely a hermeneutic or interpretative idiom and a biological or hydraulic idiom. I used this contrast to dramatise a contrast between Glock's hermeneutic ethologist who champions the capacity approach, and my hydraulic ethologist who favours the teleosemantic capacity approach. By using this contrast, I argued that the teleosemantic capacity approach does not hold that the mind is a sort of thing or stuff, but rather that the mind is defined by an animal's capacity to produce mental representations. Moreover, the hydraulic ethologist escapes the *virtus-dormitiva* objection, while the hermeneutic ethologist does not. I also argued that Glock's master argument for fact perception in animals can eliminate neither the new object objection nor the behaviourist tale without having recourse to some mental machinery. The claim that animals see objects and how they see objects must be grounded in representational mechanisms such as perceptual constancies and opponent processing. Finally, I have shown how the teleosemantic capacity approach of the hydraulic ethologist clears the field from false dichotomies and thereby creates fertile ground for Glock's arguments for forms of rationality in the animal kingdom.

Obviously, the hydraulic ethologist cannot do without behavioural data: Hermeneutic ethology without hydraulic ethology is empty, hydraulic ethology without hermeneutic ethology is blind. For what animals do and what they are capable of is the starting point for understanding animal minds. But we do not understand what animals can perceive and how they perceive it without taking into account more than mere behavioural data. What we need is knowledge about the mental machinery producing mental representations of the animal's environment, because that is where mind starts. Ernest Gellner thought that in the end Freud had to use both vocabularies for his theory of the human psyche, and also that he succeeded in uniting them (I, for one, am rather sceptical about the last claim). But the way Gellner describes the unification of the idioms seems to reflect the handshake between the hermeneutic and the hydraulic ethologists:

> A purely hermeneutic psychoanalysis would not sound like science, confer no power, and few men would turn to it in distress; a purely physicalist or biological psychoanalysis would have been too much like a science, and no fun. But the plausible-sounding fusion of both is very different, and most attractive.
>
> (Gellner 1993, 212–213)

Notes

1 Dennett characterises Cartesian materialism as

> the view you arrive at when you discard Descartes' dualism but fail to discard the imagery of a central (but material) Theatre where 'it all comes

> together' [...]. [It] is the view that there is a crucial finish line or boundary somewhere in the brain, marking a place where the order of arrival equals the order of 'presentation' in experience because what happens there is what you are conscious of.
>
> (Dennett 1991, 107)

2 'Indeed, they seem to conceive of man in Nature as a dominion within a dominion. For they believe that man disturbs, rather than follows, the order of Nature. [But] nothing happens in Nature which can be attributed to any defect in it, for Nature is always the same, and its virtue and power of acting are everywhere one and the same, that is, the laws and rules of Nature, according to which all things happen, and change from one form to another, are always and everywhere the same' (Spinoza 1985, 491, *Ethics* III. Pref.). As Michael Della Rocca explains:

> Spinoza's own view is one according to which [...] human beings and all else operate according to the same laws. Such a unification of explanatory principles is the heart of Spinoza's naturalism about psychology: human psychology is governed by the same principles that govern rocks and tables and dogs.
>
> (Della Rocca 2007, 852)

While the inclusion of rocks and tables into biological evolution certainly overshoots the mark, the dog fits the claim that both human and non-human animals are biological machines and are collections of special-purpose and general-purpose mechanisms for self-preservation.

3 In her contribution to this volume, Maria Alvarez formulates a comparable objection to Glock's argument (cf. p. 196). However, Alvarez chooses a different description of the object that the dog perceives, namely 'the dog sees such and such so arranged'. Alvarez's proposal has the advantage of being ontologically parsimonious because, unlike my new object objection, it does not need to introduce new objects into the dog's world (such as 'bowly-bone'). Another proposal would be that the dog sees ordinary objects like bones and bowls in different spatial relations. This second proposal would also be ontologically more parsimonious and closer to actual research in cognitive ethology. Finally, there is the possibility of describing the perceived objects as affordances. Depending on the arrangement of objects such as bones or bowls, new affordances for the dog (or for dogs) arise. Whether this proposal is ontologically more parsimonious depends on the interpretation of affordances.—I submit that all four proposals can serve the objection against Glock, which aims at stopping the inference from objects to facts. I, for one, see no pressure to apply ontological parsimony to the world of dogs, cats, bats, eagles, sharks, octopus, hoverflies, and other animal species. The delineation of the world into objects, properties, and kinds may very well differ from species to species. Since perception is a product of evolution, the perceptual systems have been shaped by natural selection. Following Donald D. Hoffman, there is some plausibility to the almost Nietzschean claim that veridical perceptions—systems tuned to the real structure of the world—are dominated by nonveridical systems tuned to fitness. Cf. the so-called interface theory of perception (Hoffman et al. 2015).

4 A cat that is the roommate and companion of a dog not only lives with the dog under the same roof, but they also form a community of mutual benefit. In urban slang, a person 'in da house' is a knowledgeable and helpful person in all things considered difficult.

5 This argument goes back to the distinction between practical object recognition and theoretical object recognition introduced by Ruth Millikan.

Millikan thinks that this distinction mirrors an important difference between animal minds and human minds. According to Millikan, facts as such are immensely important for human beings:

> We humans, on the other hand, collect and remember facts of kinds for which neither we nor our ancestors have yet found any practical uses. We are capable of learning thousands of facts about what has occurred or is occurring at times and in places to which we have no potential access, let alone past or present practical acquaintance. The non-fiction sections of libraries are repositories, largely, for immense collections of such facts.
> (Millikan 2004, 213)

6 This argument, of course, goes back to Gareth Evans' so-called generality constraint:

> It seems to me that there must be a sense in which thoughts are structured. The thought that John is happy has something in common with the thought that Harry is happy, and something in common with the thought that John is sad.
> (Evans 1982, 100, cf. 104)

7 One way to understand visual capabilities in animals is to assess perception of geometric illusions. There are, however, some complications with dogs (Byosiere et al. 2018).

8 What about the alternatives mentioned in n. 3 (p. 182)? Take Alvarez's proposal that the dog sees objects (such as bones and bowls) so arranged. This is very close to the proposal that the dog sees objects in different spatial relations. Hydraulic ethology suggests that the dog tracks material objects such as bones or bowls, and not complex new objects (such as 'tabley-bone' or 'bowl-bone'). Let us say that the arrangement 'bone-close-to-bowl' affords grabbing while the arrangement 'bowl-close-to-table' affords waiting. The dog's behaviour (grabbing, waiting) does not presuppose that the dog perceives facts—or so the objection to Glock's argument goes. How can hydraulic ethology help Glock out of this predicament? That is simple. The dog will grab the bone in numerous other arrangements, too. (It is easy to test.) The only thing that remains stable is the bone in different arrangements. Therefore, it is only the fact that there is a bone in front of the dog combined with the conditions that there is no learned or natural obstacle that prevents the dog from grabbing the bone now that explains the dog's behaviour. The relevant perceived fact is that the bone is in this position in front of the dog. In other words, the dog sees the fact that there is a bone in such and such a position in front of him. Glock does not need the fact that the bone is in the bowl for the explanation of this piece of dog behaviour. The behaviour is explained by the fact that there is a bone in such and such a position in front of the dog *plus* the desire for a (or this) bone *plus* the absence of natural or learned obstacles. In the case of the bone on the table, the perceived fact that there is a bone in such and such a position in front of the dog *plus* the desire for a (or this) bone *plus* the presence of natural or learned obstacles (in this case the obstacle based on associative learning, 'Never reach for anything in such and such spatial arrangement to a table') is sufficient. Not all reasons are facts.

References

Boyle, M. (2016). Additive Theories of Rationality: A Critique. *European Journal of Philosophy* 24: 527–555.

Burge, T. (2010). *Origins of Objectivity*. Oxford: Clarendon Press.
Burge, T. (2011). Origins of Perception. *Disputatio* 4: 1–38.
Burge, T. (2014). Perception. Where Mind Begins. *Philosophy* 89: 385–440.
Byosiere, S.-E. et al. (2018). Do Dogs Demonstrate Susceptibility to a Vertically Presented Ponzo Illusion? *Animal Behavior and Cognition* 5: 254–267.
Correia-Caeiro, C., Guo, K., and Mills, D.S. (2020). Perception of Dynamic Facial Expressions of Emotion between Dogs and Humans. *Animal Cognition* 23: 465–476.
Davidson, D. (1987). Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association* 60: 441–458.
Della Rocca, M. (2007). Spinoza and the Metaphysics of Scepticism. *Mind* 116: 851–874.
Dennett, D.C. (1991). *Consciousness Explained*. Boston, MA: Little Brown.
Dretske, F.I. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.
Evans, G. (1982). *Varieties of Reference*. Oxford: Oxford University Press.
Feng, L.C., Chouinard, P.A., Howell, T.J., and Bennett, P.C. (2017). Why Do Animals Differ in Their Susceptibility to Geometrical Illusions? *Psychonomic Bulletin & Review* 24: 262–276.
Fodor, J. and Lepore, E. (2001). Brandom's Burdens: Compositionality and Inferentialism. *Philosophy and Phenomenological Research* 63: 465–481.
Gellner, E. (1993). *The Psychoanalytic Movement. The Cunning of Unreason*. London. Blackwell.
Glock, H.-J. (2009). Can Animals Act for Reasons? *Inquiry* 52: 232–255.
Glock, H.-J. (2010). Can Animals Judge? *Dialectica* 64: 11–33.
Glock, H.-J. (2013). Animal Minds: A Non-Representationalist Approach. *American Philosophical Quarterly* 50: 213–232.
Glock, H.-J. (2019a). Agency, Intelligence and Reasons in Animals. *Philosophy* 94: 645–671.
Glock, H.-J. (2019b). Aristotle on the Anthropological Difference and Animal Minds. In Keil, G. (ed.), *Aristotle's Anthropology*. Cambridge: Cambridge University Press, pp. 140–160.
Glock, H.-J. (2020). Determinacy of Content: The Hard Problem about Animal Intentionality. *Harvard Review of Philosophy* 27: 101–120.
Hoffman, D.D., Singh, M., and Prakash, C. (2015). The Interface Theory of Perception. *Psychonomic Bulletin & Review* 22: 1480–1506.
Leibniz, G.W. (1991). *Monadology*. Edited by N. Rescher. Pittsburgh: University of Pittsburgh Press.
Macdonald, G. (2006). The Two Natures: Another Dogma? In Macdonald, C., and Macdonald, G. (eds.), *McDowell and His Critics*, London: Blackwell, pp. 222–235.
Matthen, M. (1999). The Disunity of Color. *Philosophical Review* 108: 47–84.
Matthen, M. (2018). Novel Colours in Animal Perception. In Andrews, K. and Beck, J. (eds.), *Routledge Handbook on Animal Cognition*. London: Routledge, pp. 65–75.
McDowell, J. (1994). *Mind and World*. Harvard, MA: Harvard University Press.
McDowell, J. (2002). Knowledge and the Internal Revisited. *Philosophy and Phenomenological Research* 64: 97–105.

Millikan, R. (2004). *Varieties of Meaning*. Cambridge, MA: MIT Press.
Neander, K. (2017). *A Mark of the Mental. In Defense of Informational Teleosemantics*. Cambridge, MA: MIT Press.
Nietzsche, F. (2002). *Beyond Good and Evil*. Edited by R.-P. Hostmann and J. Norman. Cambridge: Cambridge University Press.
Schulte, P. (2021). The Nature of Perceptual Constancies. *Philosophy and Phenomenological Research* 103: 3–20.
Spinoza, B. (1985). *Ethics* (The Collected Writings of Spinoza. Vol. 1 by Edwin Curley). Princeton, NJ: Princeton University Press.
Thompson, E. (1992). Novel Colors. *Philosophical Studies* 68: 321–349.

10 Intelligence and reasons in animals

Maria Alvarez

Introduction and preliminaries

Philosophical interest in non-human animals ('animals', from now on) is not new—Aristotle, for example, wrote widely about them, their natures and capacities from both a biological and a philosophical point of view, and did so out of intrinsic interest in them and also as an illuminating foil to corresponding questions about humans. Gradually, however, philosophers appeared to lose sight of the Aristotelian insights arising from his contention that humans *are* animals. That change, as well as encouraging doubtful conceptions of the human 'self' or 'person', developed by Descartes for example, resulted also in implausibly sceptical views about the mental capacities of animals, not just concerning their rationality or intelligence but also their capacities for perception or sensation.

Contemporary philosophy has witnessed a resurgence of interest in animals with a wider focus than before, shaped partly by the impact of cross-disciplinary approaches. Rapidly growing scientific knowledge about humans and other animals, together with changing moral sensibilities, has generated more nuanced and less assured philosophical views than in the past. Philosophers continue to debate traditional questions, such as similarities and differences between the capacities for perception, sensation, purposive agency, or intelligence in human and non-human animals; or the extent to which animals are capable of thought, rationality, and even morality. But these discussions are more tightly informed by advances in biology, anthropology, psychology, neuroscience, ethology, and evolutionary science.

Hanjo Glock has been an important participant in these philosophical debates about animals. His contributions are characterised by their ambitious scope—he tackles hard questions, such as the nature of life, intelligence, or animal minds; by a judicious use of empirical work to frame philosophical problems, and by careful attempts to identify the contribution that different disciplines can make to each question. And I have learned a lot about these issues from reading Glock's work. Of the various problems relating to animals that Glock has examined extensively and illuminatingly, I shall focus here on the question whether

DOI: 10.4324/9781003202929-14

some animals can act for reasons because, I think, it requires us to reflect on the nature of the capacity to act for reasons and its significance for humans.

I find much to agree with in Glock's arguments about animal minds and intelligence. Here, though, I shall focus on his arguments about how to assess the question whether animals can act for reasons. First, Glock has argued that a current philosophical tendency to 'move from a "subjectivist" to an "objectivist" conception of reasons[1] eliminates a substantial obstacle to the idea that animals can act for reasons'. According to him, 'instead of requiring a second-order awareness of one's own mental states, objectivism only requires a capacity to act in the light of non-mental states of affairs'(Glock 2019, 647).

Despite my commitment to an 'objectivist' conception of reasons for action, I want to suggest a different interpretation of the significance of the apparently competing conceptions of reasons. I shall argue that there is a way to interpret the 'subjectivist' view of reasons that does not imply that a creature must have second-order awareness of its own mental states in order to act for reasons. On that interpretation, neither conception presents the obstacle Glock identifies.

Second, despite various caveats and qualifications, Glock maintains that animals can act *in light of* reasons and also *reflect on* those reasons (see Glock 2019, 646). Here, I shall put pressure on one of Glock's central arguments for the view that animals act for reasons. My attempt to undermine that argument, if successful, will not, of course, show that animals cannot act for reasons. There are issues I will not address concerning his argument, different arguments, and examples of other animal behaviour that may support the attribution of the capacity to animals. However, I hope that my discussion will help to highlight what we attribute to animals when we attribute to them the capacity to act for reasons.

As Glock emphasises, answers to questions about animal minds depend 'not just on empirical findings [...] and scientific theories, but also on what one makes of heavily contested concepts' (Glock 2017, 327). Concerning the question that guides this chapter, the most significant contested concepts are those of *reasons* and *acting for reasons*; although, as we shall see, the concepts of *perception* and *cognition* are also relevant. Moreover, in this discussion, I shall follow Glock in deploying 'Morgan's Canon':

> In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale.
> (Morgan 1984, 53)[2]

The interpretation and usefulness of Morgan's Canon has been much debated, in part because Morgan's talk of 'higher' and 'lower' 'psychical

faculties' or 'psychological processes' calls for an elucidation and justification of the alleged hierarchy.

While a systematic application of the Canon may require a general account of that hierarchy, that is not necessary in order to articulate my disagreement with Glock. For both Glock and I agree that acting for reasons involves acting guided by how things are or appear to the agent to be. That requires the capacity to know or believe that 'things are thus and so', for example, that the bone is in the bowl, or that the cat is up the tree. Our disagreement is whether the behaviour of some animals needs to be explained by attributing to them those capacities for knowledge and belief rather than some less psychologically demanding perceptual and cognitive capacities. The comparative judgement here depends not on how complex or numerous the relevant psychological processes might be but on degrees of cognitive sophistication: whether the relevant cognitive capacities involve, for example, conceptualisation, abstraction, predication, grasp of logical structure or inferential relations, etc.

Reasons for acting and acting for reasons

What is a reason for acting? In contemporary philosophy, this question is often answered by first drawing a distinction between different kinds of reasons for acting (or 'practical reasons'). Typically, and simplifying somewhat, two basic kinds of practical reason are distinguished:

- 'normative reasons': reasons that *there are for A to do something*; and
- 'motivating or explanatory reasons': the reasons *for which A does something*.

The distinction is intuitive and plausible: Normative reasons are said to 'favour' your doing something, while motivating/explanatory reasons guide you in acting and thus explain why you so acted.

The notion of a reason *favouring* someone's acting is, as I construe it, the idea that a reason can make an agent's doing something right or good—instrumentally or intrinsically—along some dimension of evaluation: prudential, aesthetic, moral, hedonic, and so on.[3] The notion of a reason that *guides* an agent is, again roughly, the idea that a reason constitutes (part of) the grounds on which an agent acts. The agent is aware of a reason that is a premise in the agent's explicit or implicit practical reasoning. Such practical reasoning may be means-ends (that doing A is a means to do B), or constitutive (doing A *constitutes* doing B, which is the agent's goal). Or, alternatively, it could be more broadly deliberative: reasoning that includes the assessment of potential goals, as well as of means; or justificatory: reasoning about what makes so acting good, valuable, or somehow worthwhile.[4]

When an agent acts motivated by a reason, the reason appears to the agent to *favour* her action, that is, it appears to the agent to make her so acting right or good, at least 'pro-tanto'—that is, relative to that reason. For example, the fact that they will protect his eyes is a reason for Tim to use sunglasses. That is a (prudential) normative reason for Tim to use sunglasses. If Tim is guided to use sunglasses by the fact that they will protect his eyes, then Tim is guided by a motivating reason that is also a normative reason for him to act.[5]

The standard classification takes account of the fact that reasons can favour and/or motivate actions. But reasons can also explain actions and, crucially, a reason may (psychologically) explain someone's action without being the reason that motivated him to so act.[6] Suppose that John breaks his friendship with Peter because he finds out that Peter has betrayed him. The fact that John *knows* that Peter has betrayed him explains John's action. Call that an 'explanatory' reason. But the reason for which John breaks the friendship, in the sense of what *guides* him to so act, is not the fact that he knows something, but rather the thing he knows, namely that Peter has betrayed him. In this case, then, we have two different, though related, reasons: that Peter has betrayed John and that John knows that Peter has betrayed him. These play different roles: The first (the fact of the betrayal) motivates John to break his friendship with Peter; and the second (the fact that John knows about the betrayal) explains why he does it. Thus, if classification of reasons reflects the roles they play, we should classify motivating and explanatory reasons as two different kinds, since they play different roles—even if the same reason can play both roles. After all, the same is true of normative and motivating reasons.[7]

Accordingly, I propose a three-fold classification of practical reasons:

- 'normative reasons': reasons that *there are for A to do something*;
- 'motivating reasons': reasons that *guide A in doing something*, and
- 'explanatory reasons': reasons that *explain why A does something*.[8]

We need not take this classification to imply an ontological distinction between these kinds of reasons. This classification allows that reasons may be facts, *or* what is or what merely can be the case; *or* actual or merely possible states of affairs, *or* true or false propositions, or a combination of those. Elsewhere, I have defended the claim that *all* reasons are facts (Alvarez 2010), but for the purposes of my argument here, that claim can be put aside: What matters is that reasons are expressed by 'that' clauses. Since a reason can motivate an agent only if the agent is aware of it, for a reason to motivate an agent, the agent must know or believe what the relevant 'that' clause expresses: that the cat is up the tree, that tomorrow will rain, that there is food in the bowl, or whatever.

This three-way distinction is important here because the question whether animals can act for reasons is a question about whether animals can act *guided* by reasons. It is not a question about whether there are reasons that favour their acting in certain ways, or about whether their actions can be explained by citing some reasons. In other words, the question at issue is about *motivating* and not about *normative* or *explanatory* reasons. This distinction will be important in my assessment of Glock's arguments about competing conceptions of reasons for which someone acts and in assessing what is involved in attributing to animals the capacity to act for reasons.

Glock on 'subjectivist' and 'objectivist' conceptions of reasons

In several of his papers on this topic, Glock presents what he describes as the 'two opposing conceptions of motivating reasons', which he labels 'subjectivist' and 'objectivist', respectively. He outlines their core respective claims as follows (Glock 2019, 664):[9]

> (S) The reasons for which A acts are mental states of A (beliefs and 'pro-attitudes' like desires and intentions); and
> (O) The reasons for which A acts are facts (as A sees them) or states of affairs.

Glock goes on to argue that, if 'reasons for action were subjective states of agents, acting in their light would require A to entertain thoughts about A's own thoughts' (Glock 2019, 664). Glock's thought is that, if (S) characterises *motivating* reasons, then acting guided for that sort of motivating reason requires thoughts about one's thoughts (second-order thoughts, or metacognition) because, as noted above, for a reason to motivate an agent, the agent needs to be aware of the reason. And on this construal of (S), motivating reasons are facts, etc., about one's mental states.[10]

An objector may argue that an agent can act motivated by awareness of its mental states without the capacity to have second-order thoughts: thoughts about those mental states. For example, a hungry horse may eat because it feels hungry without self-ascribing the corresponding mental state—that is, without having the thought 'I feel hungry'. For, the objection goes, the horse is motivated to eat by its awareness of its hunger and, since hunger is a mental state, the horse is motivated by its mental state without the corresponding second-order thought.

The objection fails, for two related reasons. First, while feeling hungry is a way of being conscious because it involves experiencing certain hunger-associated sensations, it does not follow that being hungry is

being conscious *of the fact* that one feels hungry, for being conscious of that fact requires the capacity to have first-personal thoughts such as 'I am hungry'. Second, as explained in the section 'Reasons for acting and acting for reasons', motivating reasons guide by figuring in the agent's implicit or explicit practical reasoning. Felt hunger is a state that (normally) motivates the hungry creature to act—typically, to seek and consume food. This makes hunger a motivating psychological (in so far as it is felt) *state*. But that does not make hunger into a motivating *reason* precisely because, as the objection notes, hunger can motivate action while wholly bypassing any explicit or implicit reasoning process. And this is true, at least in principle, for humans too. Therefore, being motivated to act by feelings of hunger is not being guided by a mental state that motivates action *as a* reason.

An effective objection to Glock's claim here would need to show that it is possible for a mental state to be a motivating *reason* (and not just a motivating state) without the need for the agent to have (implicit or explicit) thoughts about the mental state.

This means that if (S) is taken to characterise motivating reasons, then, as Glock argues, commitment to (S) raises the bar for the capacities required to act guided by reasons—and it raises it well above those capacities that many think it is plausible to attribute to animals.[11] By contrast, Glock says, the 'objectivist' conception does not face this obstacle because it merely requires that agents should have the capacity to have beliefs, or to know facts, about the world and not about themselves. On the objectivist conception, 'reasons for action are things that are believed or known rather than subjective states of believing', which 'removes the threat posed to the idea of animals acting in the light of reasons'; Glock goes on:

> Animals can adjust their actions with a view to how things are or appear to them to be. Consequently, they are capable of acting in the light of facts or states of affairs, in the light of how things are in their environment or of how they believe them to be. This is just a corollary of their cognitive capacities. Animals can perceive and hence believe and know that things are thus-and-so.
>
> (Glock 2019, 664)

I agree with Glock that (O) is the more plausible conception of motivating reasons. However, I want to suggest an alternative way of setting up the dialectic about reasons for action that I think is a better way of understanding what at least most authors who defend (S) *really* think about practical reasons—although, admittedly, it is not what they say explicitly. I do not mean to imply that Glock is wrong in interpreting the debate as he does, given the way many participants present their views. My

claim is that the lack of an explicit distinction between motivating and explanatory reasons has obfuscated the debate.

My suggestion is that the two conceptions of 'the reasons for which A acts' captured in (S) and (O), respectively, are not in competition because they concern different kinds of reasons, namely, explanatory (S) and motivating (O) reasons, respectively. On that construal, (S) no longer presents the obstacle to animal agency Glock identifies because it allows that the reasons that motivate agents can be facts about their environment and need not be facts about their mental states.[12]

Explanatory 'subjectivist' and motivating 'objectivist' conceptions of reasons: a reconciliation?

Glock is right in claiming that, if the reason that motivates an agent to act (in the sense of the reasons that guide her) is a mental state of hers (or facts about those mental states), acting motivated by that reason requires the sophisticated capacity to self-ascribe mental states. For example, if my desire to kill my parents is the reason that motivates me to visit a psychiatrist, I need to be aware of the desire—and, further, I must reason, say, that I need help with this appalling desire and that a psychiatrist may provide such help. A little reflection, however, reveals that most motivating reasons are not mental states in this sense.

It seems plausible then that, when philosophers have claimed that reasons for action are mental states, they had something else in mind. My suggestion is that, when endorsing claims like (S), these philosophers meant to characterise *explanatory* reasons (to use my terminology) and that they take *motivation* reasons to be the contents of those mental states. However, because they do not make the distinction explicitly but instead tend to amalgamate the two under the label 'motivating/explanatory reasons', their claims can be confusing, and sometimes confused.

When someone who endorses (S) says that the reasons for which an agent acts are mental states, it is tempting to think that they take the mental state *as such* to play both roles: explaining (by causing) and motivating (by guiding). But if we attend to the distinction between explanatory and motivating reasons, we can attribute to 'subjectivists' the view that explanatory reasons are (facts about) mental states, and motivating reasons are the contents of those mental states: propositions, or the corresponding facts, that guide their actions.

This suggestion may be better described as 'dissolving' rather than 'resolving' the objectivist/subjectivist debate. But I think there is good reason both of coherence and interpretation to do so. I shall give two illustrations. Consider first Donald Davidson, who Glock rightly casts as a representative of 'subjectivism', because he explicitly defended the claim that someone's reasons for acting are (combinations of their) mental

states (Glock 2019, 664). In the paper where Davidson spells out his view of reasons, 'Actions, Reasons and Causes', Davidson is concerned with a specific question, which appears in the opening of the paper:

> What is the relation between a reason and an action when the reason *explains* the action by giving the agent's reason for doing what he did?
> (Davidson 1963, 685, my italics)

Davidson's question is about reasons that *explain* an action—in my terminology, it is a question about explanatory reasons. And his answer is that the relation between the reason that explains an action ('by giving the agent's reason for doing what he did') and the action is causal: The reason *causes* or, to be more precise, is a *causal condition* of the action.

Davidson held, roughly, that a reason that explains an intentional action is a combination of two mental states (a belief and a 'pro-attitude') that (i) causally explain the action because they are a causal condition for its occurrence;[13] and (ii) *rationalise* the action under some description because, roughly, their contents enable us to see what about that action so described appealed to the agent (Davidson 1963, 693ff). In my terminology, we can say that, for Davidson, the combination of the mental states is an *explanatory* reason because it explains the occurrence of the action; and the contents of those mental states are the agent's *motivating* reason: They guide the agent in acting through her practical reasoning.[14]

And, as a second illustration, here is what Stephen Darwall says about the apparent conflict between the 'subjectivist' view in Smith (1994) and Dancy's (2000) 'objectivism':

> 'Motivating reason' in Dancy's pen means the agent's reason, the (believed, putative) fact in light of which the agent acted. Smith, however, uses 'the agent's normative reason' to refer to this and 'motivating reason' to refer to the desire/belief combination necessary to explain behavior teleologically.
> (Darwall 2003, 442–443)

In short, if the subjectivist claim (S) is construed in the way I have suggested, as combining subjectivism about explanatory reasons with objectivism about motivating reasons, then it is no longer clear that (S) places a higher requirement than (O) concerning the capacities an agent must have in order to act for reasons: Both now require merely that agents have desires and beliefs (or knowledge) whose contents can guide their possessors in acting.

I now turn to Glock's contention that the behaviour some animals display, together with the fact that they can perceive their environment, suffices to show that some animals act guided by reasons.

Perception, cognition, and acting for reasons

Glock objects to what he calls the 'lingualist' family of arguments against the view that animals act for reasons.[15] One version of the argument is, roughly, that the capacity to grasp and reflect on reasons requires linguistic capacities that (most) animals lack. It follows that these animals lack the capacity to grasp and reflect on reasons. Against this *a priori* argument, Glock claims that the empirical evidence shows that some animals who lack those linguistic capacities are capable of behaviour whose complexity and intelligence can only be explained by attributing to them the capacity to grasp and reflect on reasons. Therefore, and *contra* lingualism, the capacity to act for reasons cannot require such linguistic capacities. I will not comment on Glock's criticism of 'lingualism' here, but will instead focus on a positive argument of Glock's to defend the claim at issue, which relies on the capacity animals have to perceive their environment.

As cited in the section 'Glock on "subjectivist" and "objectivist" conceptions of reasons', Glock says that the claim that animals are 'capable of acting in the light of facts or states of affairs' is 'just a corollary of their cognitive capacities'; animals, he adds, 'can perceive and hence believe and know that things are thus-and-so' (Glock 2019, 664). But this seems too quick. It is undeniable that some animals perceive and represent their environment and, as Glock argues persuasively in various places, those perceptual-cognitive capacities are essential to explaining the highly sophisticated forms of behaviour that these animals display—behaviour that evinces intelligence: the problem-solving ability that involves a grasp of situations and the capacity to apply this understanding in novel situations in pursuit of their goals.[16] The question, however, is what follows from this about whether animals act for reasons in the sense at issue: that is, what follows about whether they are *guided* by reasons.

If an animal can perceive 'that things are thus-and-so', then it can know, and perhaps believe, that things are thus-and-so. But of course, perceiving (seeing, hearing, smelling) is not necessarily perceiving (seeing, hearing, smelling) *that* things are thus-and-so. Therefore, whether a creature with perceptual capacities can have factual knowledge of its environment depends on *what* the creature perceives when it perceives its environment. It is a familiar thought that perception can have different types of objects: things (such as cats, trees, or water), the colours, smells, textures, sounds of those things, happenings, such as events and actions (the cat's running up the tree) or facts (that the cat is up the tree). And, in principle, it is possible to perceive things, their colours, etc. without thereby perceiving or otherwise knowing the corresponding facts. For example, it is possible to see a cat without seeing or otherwise knowing that the thing seen is a cat; to see the black colour of the cat without

seeing or otherwise knowing that that is the colour of the cat; to hear the cat's running up the tree without hearing or otherwise knowing that the cat is running up the tree. By contrast, to see that the cat is up the tree *is* a way of knowing that the cat is up the tree.[17]

This simply reminds us that it is possible to perceive a thing that is F or that is Φ-ing without knowing that the thing perceived is F or that it is Φ-ing, i.e. without perceiving the corresponding fact. A crucial question, for us, then, is whether it is possible for a creature to perceive its environment—things, their properties, happenings, etc.—without thereby having any *factual* knowledge about its environment.[18]

This question should not be conflated with the question whether it is possible for a creature to perceive things, their properties, occurrences, etc. without thereby having any knowledge *of* what it perceives. If perception is a form of awareness of whatever is perceived, being aware of a thing is having some form of knowledge, some cognition, *of* that thing. Given this, although the possession of perceptual capacities implies some capacity for cognition, the question at issue is whether the possession of perceptual capacities implies the capacity for '*fact*-cognition', that is, cognition that things are thus-and-so. This is a more general and difficult question that likely requires a combination of empirical evidence and philosophical argument to settle.

In principle, it seems possible that a creature should, through its perceptual capacities, know things, their properties, events, processes, etc. involving those things, while lacking any fact-cognition. This knowledge may require that the creature should have mental representations but, if so, these need not amount to fact-cognition.[19] For example, it seems plausible that a human baby may know its mother without yet having the ability to have any fact-thoughts about her or about anything. The baby may know, i.e. be able to recognise, its mother by her voice, her smell, her face, etc. without having any capacity for fact-thoughts such as 'this is my mother's voice' or 'this is my mother's smell' or any others. The baby's knowledge is evinced in, or perhaps just is, its capacity to recognise its mother, to discriminate between her and other humans, creatures, or things, as demonstrated in its emotional reactions, behaviour, etc. There is no reason to think that the baby's knowledge of its mother requires or involves fact-perception or fact-cognition of anything. And this suggestion seems to extend to the baby's capacity to perceive and so have some cognition of other things such as the colours of things, actions, processes, etc. For example, the baby may react with fright to hearing angry shouting, with enjoyment to playful interaction; it may respond to a game that depends on perceiving changes, e.g. an object's being first visible and then concealed, etc. These possibilities suggest examples of perceptual cognition of those various things that explain the baby's behaviour without the need to credit the baby with the capacity for fact-cognition. Of course, many animals are capable of much more complex

behaviour than human babies. Nonetheless, if this is possible, whether animals are capable of fact-cognition must be established independently of the general attribution to them of the capacity to perceive their environment and to exploit this capacity in adapting their behaviour flexibly in pursuit of their goals.

One way to establish whether animals can perceive and know facts, is to examine whether the behaviour they exhibit requires ascribing to them an awareness of *facts*, as opposed to mere perceptual awareness of things, colours, sounds, occurrences, etc. (which, for ease, I shall call 'object-perception'). This is precisely what Glock claims to do in this passage:

> Consider a dog that has learned not to grab anything when it is lying on the table but only when it is lying in his bowl. This dog now sees a bone on the table, but refrains from grabbing it and instead looks on, panting. Yet as soon as the bone is placed in the bowl, the dog goes for it. This mundane sequence of events is not explained by the dog simply perceiving discrete objects: the bone, the table, and the bowl. It can only be explained in terms of the following opposition:
>
> - The dog sees at time t_1 that the bone is on the table.
> - The dog sees at time t_2 that the bone is in the bowl.
>
> Why? Because at both t_1 and t_2 the dog can see bone, table, and bowl. So perception of the conglomeration formed by these three objects cannot explain the difference in its reactions at t_1 and t_2. One might respond that the problem vanishes if spatial relations like *x being on y* are among the objects that the dog can perceive. However, simply perceiving three distinct objects—bone, table, *x being on y*—does not explain the dog's behavior. Such an explanation is only in the offing if the dog also can perceive that the bone stands in the relation of *being on* to the table at one moment, to the floor at the next. And in that case we are back with perceiving that *p*.
>
> (Glock 2017, 338)

Glock identifies two explanatory options for the dog's behaviour: object-perception (and cognition) or fact-perception (and cognition). He deems the first unsatisfactory because insufficient to explain the difference in behaviour. The second *is* satisfactory but it credits the dog with fact-cognition. And, in so far as acting for reasons is acting guided by (putative) facts, this allows that the dog acted guided by the *reason* (fact) that the bone was in the bowl.

I will raise two objections to Glock. The first is that he does not say enough to undermine the explanation by object-perception. The second is that he underestimates the complexity of the psychological capacities we need to attribute to animals on his preferred explanation.

I agree with Glock on one thing. The fact that the dog takes the bone at t_2 but not at t_1 is explained by the fact that the bone is in the bowl at t_2 but not at t_1. We disagree about what type of psychological cognitive processes in the dog underlie this explanation. He claims that the dog needs to *know* the fact about the bone, whereas I want to suggest that the dog's perception of the objects at issue—table, bone, and bowl—suffices to explain the dog's action.[20]

Glock objects that postulating mere perception of the relevant objects does not suffice, because 'at both t_1 and t_2 the dog can see bone, table, and bowl'. He notes that his opponent may respond that what the dog 'perceives is *bone on table* or *bone in bowl*,' but Glock counters that this would explain the different behaviour, given perception of the *same* bone, only if these descriptions of what the dog sees are 'used as ellipses for [*bone*] *lying on the table* and [*bone*] *lying in the bowl*'. And, he adds,

> to perceive the bone *as* lying in the bowl is to perceive—albeit by another name—that the bone is lying in the bowl. One way or another, the dog's behavior can be explained only on the assumption of factual perception, perception *that*.
>
> (Glock 2017, 339)

The argument is not convincing.[21] Even if we concede that the dog's perceiving the bone lying in the bowl is, as Glock claims, its perceiving the bone *as* lying in the bowl, it does not follow that the latter is also its perceiving *that* the bone is lying in the bowl. That equation is controversial.[22] Be that as it may, Glock's opponent may find alternative ways of explaining how perception of objects suffices to explain the dog's behaviour which make no reference to the concept of 'perceiving as'. Here is a suggestion.

While it is true that, as Glock claims, the dog can perceive the three objects at both times, at each time, it perceives the same objects *differently arranged*, so that the world will look, smell, etc. quite different to the dog at t_1 and at t_2. It is not merely that the dog perceives the same three objects, which, at each time, happen to be differently arranged. Rather, the dog perceives the-same-three-objects-differently-arranged—or, as we might put it, it perceives two *different arrangements* of the same objects at each time.[23]

Glock will object that talk of the capacity to perceive and discriminate between different *arrangements of objects* is just a way of smuggling in the idea of perceiving *facts* about those objects. That would be so, however, only if an arrangement of objects is some abstract thing over and above the objects so arranged. I cannot see that it need be. The dog's capacity to perceive and discriminate between arrangements of objects surely depends on the fact that the word looks, smells, etc. quite different to the dog at each time precisely because the objects are differently arranged at each time. And, if the dog perceives those two different arrangements in succession,

this difference in the way the world looks, smells, etc. to the dog, together with its training and its doggy liking of bones, seems sufficient to explain why the dog grabs the bone only when it is lying in his bowl.[24]

Much more would need to be said to flesh out this suggestion. But if something like it is, in principle, plausible, then, we have an account of why the fact that the bone is in the bowl at t_2 explains why the dog takes the bone at t_2 and not before, without attributing to the dog the capacity to be guided by facts—i.e. to act for reasons. I do not think Glock says enough to eliminate this possibility.

I now turn to the second part of my objection—that Glock underestimates the complexity of the psychological capacities that we need to attribute to animals on his preferred explanation.

As we saw above, motivating reasons are reasons that seem to the agent to favour his or her action. This implies that facts explain actions *as an agent's reasons* only if the agent is credited with the capacity for some sort of normative or axiological thinking: the capacity to grasp the thought that certain facts make a certain course of action right or good for one. Because, for example, the fact that the bone is now in the bowl explains why the dog takes it now *as the dog's reason* for taking it now only if we ascribe to the dog, not just the thought that the bone is in the bowl, but *also* something like the thought that *the bone's being in the bowl somehow favours*—makes it right or good or at least ok—for it [the dog] to take the bone. Without attributing something like that thought to the dog, it is not clear what the claim that the fact about the bowl is *the dog's reason* for taking the bone amounts to—since motivating reasons are reasons that the agent takes to favour his or her action.

Note that the capacity for normative or axiological thoughts that I am claiming is required for acting for reasons is different from, and less cognitively demanding than, a capacity Glock acknowledges non-linguistic animals lack, namely the capacity '*to think about* one's reasoning or deliberation' (Glock 2019, 670). The normative or axiological thoughts I am concerned with are not about one's reasoning but about one's action—and they underpin choices an agent makes between different behavioural options available to her which she makes on account of facts relating to those actions that she is aware of.

Perhaps we can attribute such thoughts to animals, including the dog in our example, but we should do so only if there is no less cognitively demanding, satisfactory alternative to explain how perception of objects and arrangements thereof, or of occurrences, can guide their behaviour. One familiar suggestion is Gibson's notion of *affordance*:

> [P]laces, attached objects, objects, and substances are what are mainly perceived, together with events, which are changes of these things. To [perceive] these things is to perceive what they afford.
>
> (Gibson 1979, 240)

> The meaning or value of a thing consists of what it affords.
> (Gibson et al. 1982, 407)[25]

In our example, the bone in the bowl *is*, while the bone on the table *is not*, a positive affordance for the dog—which is to say that the former is perceived by the dog as 'eat-able', while the latter is not. That difference in perception explains the difference in the dog's behaviour between t_1 and t_2. As Andrea Scarantino explains, for Gibson 'to perceive is to pick up information for purposes of behavioural discrimination (rather than for purposes of belief fixation)' (Scarantino 2003, 953).[26] This way of understanding how perception of objects and their arrangements *can* explain the behaviour under discussion is, in fact, in line with Glock's own qualification to his claim that animals can act guided by, and evaluate, facts: 'or, in a more realistic ethological idiom, [they can evaluate] features of their environment—*in the sense of responding to them as good or bad (e.g. attractive or threatening*)' (Glock 2019, 670, my italics), and act accordingly. In short, animals may be able to evaluate objects and situations, in the sense that they can perceive them as 'having certain affective, conative or emotional dimension—as dangerous, edible, interesting, appealing, etc.' or be able to recognise their 'instrumental significance' (ibid.) in the sense that they can use them for various ends, without believing or knowing *that* the object or occurrence is dangerous, edible, etc.; or that this object is a means to different ends.

Concluding remarks

I have only examined one of Glock's arguments in favour of the claim that animals can act for reasons and have tried to explain why I do not find it convincing. The claim remains controversial and in a way that leads some to wonder what, if anything, hangs on the question; to ask, for example, what follows about animal intelligence if it turns out that animals do not act for reasons in the sense I have specified.

One thing seems clear: If we reached that conclusion, it would not follow that the intelligence animals are rightly credited with, given the sorts of behaviour they display, would thereby be diminished—at least on Glock's understanding of the notion spelled out in the section 'Perception, cognition and acting for reasons'. Even if they do not act for reasons, some animals display highly intelligent behaviour in successfully pursuing goals in a flexible manner, and they do so through an awareness of their environment and the possibilities it affords them.

On the other hand, the conclusion enables us to sidestep some familiar thorny problems about precisely which beliefs, or which propositional knowledge, to attribute to animals. The problem is partly one of determinacy: which precise belief 'content' we should attribute to them.[27] It is true, as many have pointed out, that our own beliefs are sometimes

more indeterminate than we might assume, but the difference is that our beliefs can become more precise and determinate, through reasoning and reflection, and articulated in language. The same is not possible for animals. I do not mean the problem is unsurmountable—only that, if there is an alternative to understanding animal behaviour as explained by reference to 'their reasons', we not only avoid this indeterminacy problem but also have an explanation of why it seems so intractable: Their behaviour is not guided by reasons.

In my view, the main insight to be derived from the starker separation of the capacity for (highly) intelligent behaviour displayed by some animals from that of acting for reasons here suggested is that it brings into sharper focus that capacity and its role in human life. For the separation reminds us, on the one hand, that many human activities, even fairly complex ones, may be things we do without engaging our reasoning capacities, just as other animals do. On the other hand, the capacity to act for reasons, to deliberate and reflect on those reasons, and to explain and to justify our behaviour by citing them has transformed our lives as social animals and made them radically different from the lives of other animals, despite the many features shared with them. It has done so by vastly transforming the capacities for such things as communication, teaching, coordination, and cooperation that other animals also engage in. But also, and more significantly, because it has massively expanded our repertoire of capacities, enabling us to develop the extremely sophisticated range of technological, scientific, creative, artistic, and social activities that comprise human cultures and civilisations. You do not need to be a 'lingualist' to accept that those reason-involving capacities are language-dependent.

Notes

1 For a collection examining this trend especially in epistemology, see Mitova (2018).
2 See Glock (2017, 331). For a recent discussion, see Allen-Hermanson (2005).
3 This is a wider conception of normative reasons than endorsed by others. For example, Mantel seems to regard normative reasons as those that favour an action in so far as acting for that reason would have something like 'moral worth': 'Acting for a normative reason should be understood in a more demanding sense in which it reflects well on the agent' (Mantel 2018, 3).
4 See Alvarez (2010, 158ff).
5 Compare. Al sets fire to Beth's rubbish bin because Beth keeps leaving it in his front garden: He may reason that his act of arson is justified by her inconsiderate behaviour. If so, Al is motivated (in part) by *the fact* that Beth left the bin in his garden. However, that fact, arguably, does not *favour* Al's act of arson, at least not all things considered. In this case, the fact that Beth left the bin was a reason that motivated (guided) Al, although it was not a normative (justifying) reason for him to commit arson.
6 Reasons can explain other things, for instance, why something happened, or indeed why something is good or right. A normative reason explains why doing something is right but, arguably, only because it *makes* it right: The fact

that explains why the action is right does so because it is a 'right-making' feature of the action.

7 See Alvarez (2009; 2010).

8 Given the rationale for classification, I do not mean these are the only kinds of reason that might be helpful to distinguish for other purposes.

9 See also Glock (2017).

10 And, as Glock (2019, 664) notes, 'the same would hold if the reasons in the light of which A acts were psychological facts about A, e.g. the fact that A believes that p, desires X or wants to Φ'.

11 Glock lists several commentators who agree on this point, Bermúdez (2006), Hacker (2007), and Millikan (2006), while otherwise disagreeing on animal mentality.

12 For a similar suggestion, which, however, does not draw a distinction between explanatory and motivating reasons as I do, see Keil (2012).

13 Since explanation is a relation between facts, or statements of facts, Davidson's view is better expressed by saying that the reason that rationally explains the action is a combination of *facts* about those mental states. See Strawson (1992).

14 See, e.g., Davidson (1970).

15 See Glock (2017, 328ff).

16 Here, again, I follow roughly Glock's characterisation of intelligence, in, e.g. Glock (2019, sect. 3).

17 I leave aside the question whether, if we perceive facts, we do so by perceiving other things: objects, colours, etc.

18 Other related questions are whether it is possible to perceive a thing without perceiving it *as* something, and whether perceiving *as* something requires concept possession and application. But even if perception of x is always perception of x *as* F, it does not immediately follow that it is perception that x is F, i.e. that it is fact-perception. See, e.g. Dretske (1990), Mulligan (2003), and Burge (2010).

19 See, for example, Rescorla (2009; 2017).

20 Perception may suffice in this example, although in general some capacity for (non-propositional) memory and imagination will also be required in order to explain the relevant animal behaviour.

21 For similar reservations about the argument motivated by quite different concerns, see Wild, this volume.

22 See, for example, Mulligan (2003).

23 And, if the dog was attentive, it may also have perceived the *process* of the bone's being moved from the table to the bowl, as well as the event that consists in this change. That the dog can perceive that process and event might also contribute to explaining its behaviour.

24 Consider this passage from Tyler Burge:

> Some have thought that whether perception is propositional depends on whether what is represented is an object or a state of affairs. I believe that this view has things backwards. The issue is over the organization, structure, form of the representational state. A state of affairs can be represented either with *a singular representational structure* or with *a propositional structure*. Similarly, a relation between different particulars can be represented in a *nominal* or *propositional* way.
>
> (Burge 2010, 33)

25 Quoted in Scarantino (2003), which offers a helpful philosophical account of affordances, to which my remarks on this are indebted.

26 Scarantino is sceptical of the radical Gibsonian claim that 'the assimilation of information for affordances can occur without the involvement of mental representations and mental processes involving representations' (e.g., inferences, computations, retrievals of memories, etc.). (See, for example, Gibson 1979, 238–263; Scarantino 2003, 954.) Still, affordances may require mental representations or processes of memory, etc. that fall short of the cognitive capacities required for fact-perception.

27 The problem goes back to Stich (1979) and is developed in Davidson (1982). Glock discusses it in Glock (2020).

References

Allen-Hermanson, S. (2005). Morgan's Canon Revisited. *Philosophy of Science* 72: 608–631.

Alvarez, M. (2009). How Many Kinds of Reasons? *Philosophical Explorations* 12: 181–193.

Alvarez, M. (2010). *Kinds of Reasons: An Essay on the Philosophy of Action*. Oxford: Oxford University Press.

Bermúdez, J.L. (2006). Animal Reasoning and Proto-Logic. In Hurley, S. and Nudds, M. (eds.), *Rational Animals?* Oxford: Oxford University Press, pp. 127–138.

Burge, T. (2010). Origins of Perception. *Disputatio* 4: 1–38.

Dancy, J. (2000). *Practical Reality*. Oxford: Oxford University Press.

Davidson, D. (1963/1980). Actions, Reasons and Causes. Reprinted in Davidson, D. (1980), pp. 3–20.

Davidson, D. (1970/1980). How Is Weakness of the Will Possible? Reprinted in Davidson, D. (1980), pp. 21–42.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Oxford University Press.

Davidson, D. (1982). Rational Animals. *Dialectica* 36: 317–327.

Darwall, S. (2003). Desires, Reasons and Causes. *Philosophy and Phenomenological Research* 67: 435–443.

Dretske, F. (1990). Seeing, Believing, and Knowing. In Osherson, D.N., Kosslyn, S.M. and Hollerbach, J.M. (eds.), *Visual Cognition and Action* (An Invitation to Cognitive Science, Vol. 2). Cambridge, MA: MIT Press, pp. 129–48.

Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.

Gibson, J., Reed, E. and Jones, R. (1982). *Reasons for Realism: Selected Essays of James J. Gibson*. Hillsdale, NJ: L. Erlbaum.

Glock, H.-J. (2017). Animal Minds. In McLaughlin, B.P. (ed.), *Philosophy: Mind*. (Macmillan Interdisciplinary Handbooks). New York: Macmillan, pp. 327–351.

Glock, H.-J. (2019). Agency, Intelligence and Reasons in Animals. *Philosophy* 94: 645–671.

Glock, H.-J. (2020). Determinacy of Content: The Hard Problem about Animal Intentionality. *Harvard Review of Philosophy* 27: 101–120.

Hacker, P.M.S. (2007). *Human Nature*. Oxford: Blackwell.

Keil, G. (2012). Beyond Assimilationism and Differentialism: Comment on Glock. In Nida-Rümelin, J. and Özmen, E. (eds.) *Welt der Gründe*. Hamburg: Meiner, pp. 914–922.

Mantel, S. (2018). *Determined by Reasons: A Competence Account of Acting for a Normative Reason*. New York: Routledge.

Millikan, R.G. (2006). Styles of Rationality. In Hurley, S. and Nudds, M. (eds.), *Rational Animals?* Oxford: Oxford University Press, pp. 117–126.

Mitova, V. (2018). *The Factive Turn in Epistemology*. Cambridge: Cambridge University Press.

Morgan, L. (1894). *An Introduction to Comparative Psychology*. London: W. Scott, Limited.

Mulligan, K. (2003). Seeing, Certainty and Apprehension. In Fossheim, H., Mandt Larsen, T. and Rickard Sageng, J. (eds.), *Non-Conceptual Aspects of Experience* (Proceedings of the 2000 Melbu Conference on Non-Conceptual Content). Oslo: Unipub forlag.

Rescorla, M. (2009). Chrysippus' Dog as a Case Study in Non-Linguistic Cognition. In Lurz, R.W. (ed.), *The Philosophy of Animal Minds*. Cambridge: Cambridge University Press, pp. 52–71.

Rescorla, M. (2017). Maps in the Head. In Andrews, K. and Beck, J. (eds.), *The Routledge Handbook of Philosophy of Animal Minds*. New York: Routledge, pp. 34–45.

Scarantino, A. (2003). Affordances Explained. *Philosophy of Science* 70: 949–961.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.

Stich, S.P. (1979). Do Animals Have Beliefs? *Australasian Journal of Philosophy* 57: 15–28.

Strawson, P.F. (1992). Causation and Explanation. In *Analysis & Metaphysics*. Oxford: Oxford University Press, pp. 109–131.

11 A conceptual framework for empathy in humans and nonhuman animals

Albert Newen, Maja Griem, and Simone Pika

Can we develop a conceptual framework that allows us to characterise similarities and differences between cases of empathy in humans and nonhuman animals (hereafter "animals")? We suggest such a framework in three steps. First, we develop a new conceptual framework by distinguishing central components of empathy starting with a paradigmatic case of human empathy. Second, we characterise different types of empathy as they develop in human ontogenesis. Third, we describe which of these types of empathy can be found in other species based on animal studies. The integration of phylogenetic and ontogenetic perspectives enables us to outline how this conceptual framework can fruitfully be applied to explicate the relation of empathy in humans and other animals.

Introduction[1]

Let us start with a provisional characterisation of empathy that we want to develop into a scientific concept of empathy: In the *Cambridge dictionary*, we can read that empathy is "the ability to share someone else's feelings or experiences by imagining what it would be like to be in that person's situation."[2] This is a description inspired by a typical case of human empathy, e.g. the empathy of parents for their children. We argue below that such typical cases of empathy involve a registration of the other's affective state, an attitude of caring for the other, and then also a typical response behaviour. But this need not be a sharing of someone else's feeling. A typical case of empathy is also one in which I register the sadness of the other and then I comfort him with the attitude of caring. Often, empathy need not involve a sharing of feelings but can be based on having complementary affective states or processes. Furthermore, imagining what it would be like to be in someone else's situation is a quite demanding cognitive ability of high-level mental simulation, which implies that only agents who have this ability can empathise. This is not plausible since empathy can also be observed in more basic cognitive situations as we will demonstrate. Thus, we need to develop a systematic and scientifically grounded concept of empathy.

DOI: 10.4324/9781003202929-15

Why should we care about empathy? Intuitively, everyone may agree that empathy is the salt in the soup of our intersubjective life with family and friends. Even when interacting with enemies, one can accept from a rational perspective that some empathy with them seems to be relevant to overcome conflicts or to stop an escalation because it prevents a dehumanisation of enemies. Only looking at children, we have extensive scientific evidence for a multiplicity of functional roles of empathy during early education:

> The scope of functions that empathy in children can mediate include social understanding, emotional competence, prosocial and moral behaviour, compassion and caring, and regulation of aggression and other antisocial behaviours. It should be emphasised that empathy is not equivalent to these personal and interpersonal competencies, nor is it a magic elixir that automatically produces social competence and prosocial behaviour. However, it is a very important factor in the matrix of developmental variables that mediate these cognitive and affective behaviours, all of which are important to schooling.
> (Feshbach and Feshbach 2009, 86)

Furthermore, empathy can add another dimension of social interaction along which animals might act for a reason, as proposed by Glock (2019). In line with his approach, we are also looking for a middle ground that not only does not undersell, but also does not exaggerate the abilities of nonhuman animals as, e.g., such a middle ground is developed in detail for concepts (Glock 2000; Newen and Bartels 2007).

Although there is a large agreement on the relevance of the concept of empathy, there is still no consensus on how to systematically characterise the concept such that empathic processes are described adequately and we are able to distinguish them from related phenomena.

How can we methodologically reach such a proposal? *First of all*, we think that we should not aim for a definition with jointly necessary and sufficient conditions of empathy for two reasons: (a) Looking at the variety of phenomena associated with empathy and the proposals available in the literature (see below), we think that it is not very promising to continue any search for such a type of definition. (b) This type of definition is understood as the definitional understanding of concepts (Laurence and Margolis 1999, 3–81) and it only holds for concepts with very constrained features (e.g. bachelors are unmarried men) or for scientifically introduced concepts such as *gene* or *atom*. But *empathy* is a concept anchored in folk psychology and it captures such a wide range of phenomena that we propose to aim for a multidimensional characterisation of the concept such that we can discover a range of typical components and sub-features of these components, which allow us to systematically describe ideal cases of empathy and the family resemblance relations

of related phenomena to these ideal cases. *Second*, we suggest that we take a look at the best description of empathy during the development of young children because this enables us to separate relevant features that are involved in full-blown empathy. As a consequence, we may not only expect a typical case of full-blown empathy but also a sequence of idealised types of empathy that develop during ontogenesis on the way to acquire full-blown empathy. *Third*, we furthermore suggest that the framework should be rich enough to account not only for empathy in humans and during their ontogenesis but also for nonhuman animals. The latter aspect is taken as one of the important aspects that has so far been underestimated for a fruitful concept of empathy: We should account for the fact that such an ability is evolutionarily anchored and, therefore, especially try to bring together observations from ontogenetic and phylogenetic development with typical cases of full-blown empathy. One may reply that the dimension of animal empathy is too unclear to be considered as a constraint. Let us briefly summarise this challenge: Until recently, empathy has been thought of as a uniquely human ability since research investigating this ability has been strongly influenced by the Bischof-Köhler hypothesis (Bischof-Köhler 1985) postulating that the behaviour of nonhuman animals is controlled only by their own current motivational states. Hence, nonhuman animals cannot anticipate future motivational states or act on the motivational and emotional states of others. However, the past two decades of research have revealed increasing evidence of behaviour which indicates that we should at least investigate the possibility of empathy in nonhuman animals, e.g. prosocial behaviour in rodents (Bartal et al. 2011). Focusing specifically on the mechanisms underlying prosociality in rodents, studies showed overlaps of behavioural flexibility, similar behavioural dispositions (e.g. familiarity bias, sex bias with females showing more helping behaviour), as well as similar brain circuits compared to humans (Meyza et al. 2017). Many other observations lead to the criticism of the traditional human-centered view on empathy, e.g. by de Waal (2009). He argues that "being in tune with others, coordinating activities, and caring for those in need isn't restricted to our species. Human empathy has the backing of a long evolutionary history" (de Waal 2009, x of preface). Since the debate about empathy in nonhuman animals is not settled, one benefit of a conceptual framework for empathy would be that we can systematically classify different types of empathy in such a way that we can apply it to nonhuman animals.

Criteria of adequacy for a concept of empathy developed from a typical case

Tom lives with his old neighbour in Corona-times, and she tells him that she suffers from isolation and that she is sad that she is no longer able

to cook any complex meals but has to deal with simple cuisine due to her diminishing abilities. Since Tom knows her quite well, he is aware of the fact that she does not want any fancy meals that one could easily order from a restaurant, but she prefers a special cooking that has to be self-made. Tom feels sorry for her and decides to cook and bring her favourite meal for her upcoming birthday. She is delighted.

This story involves the characteristic components of full-blown empathy. The core intuition is that empathy typically involves the *registration* of the others' relevant mental states or processes ("state" for short in the following), including the affective state. This is combined with the *attitude* to take care of the other and the *selection of a supporting action* that accounts for the mental states and the situation of the person. More precisely: We think that we can cluster relevant elements into three main components, namely (1) registration, (2) attitude, and (3) behavioural response. In this example, you (1) *register and/or activate* the old lady's (1a) *affective state* of being sad, (1b) *the relevant situation* she is in, namely of not being able to cook adequately, and (1c) *her relevant mindset*, namely the desire for her favourite meals. Furthermore, Tom acts based on (2) *an attitude of care for the other.*[3] And (3) Tom decides to act such that he fulfills *her* desires (or needs), i.e. his action is typically other-directed. Thus, we want to exclude the behaviour of a cognitive system which lacks a relevant self-other distinction: Without this as a minimal basis for the response behaviour, we may at best observe a borderline case of empathy. But having a self-other distinction still allows for self-directed response behaviour, although this is not typical for full-blown empathy, e.g. if Tom would feel sorry for her but avoid any contact not to be reminded of her situation.

We take these to be the paradigmatic criteria of full-blown empathy understood as three typical components. A Wittgensteinian understanding of the concept of empathy allows us to exclude cases of non-empathy and to describe borderline cases: This understanding of empathy does not aim for a classical definition but starts out with the most typical examples as core cases of empathy, rather than a description of necessary and sufficient conditions of empathy. We suggest the following implementation of the Wittgensteinian idea of family resemblance: If we have three characteristic components of empathy, then a phenomenon that does not realise any is clearly not a case of empathy. If from three components, only one is realised, we have a borderline case. If two of three components are realised, we enter the area of fruitful descriptions of family resemblance cases. If all three components are fully realised, this is a case of full-blown empathy.

To offer a new framework for a characterisation of empathy, we need to describe how the three main components can be realised with typical sub-features: This is the basis to describe a variety of cases of empathy and enable multi-realisability. Before we develop this new account, we

critically discuss the two most influential theories of empathy that are flexible enough to account for phylogenetic evidence.

Criticizing two central conceptual frameworks

The two most promising approaches on empathy and its related phenomena are the so-called Russian doll model by de Waal and Preston (2017) and the combination model by Yamamoto (2017).

The basic pillar of the Russian doll model (de Waal and Preston 2017) follows a developmental strategy: Empathy needs a perception-action coupling mechanism as a necessary basis, and important applications of this mechanism are the abilities of motor mimicry and emotional contagion. According to the authors, all other empathic abilities are built on this core element. This application of basic forms of empathy is widened with the additional abilities of sympathetic concern and consolation. Furthermore, the most complex forms of empathy additionally involve higher-cognitive empathic abilities such as perspective taking

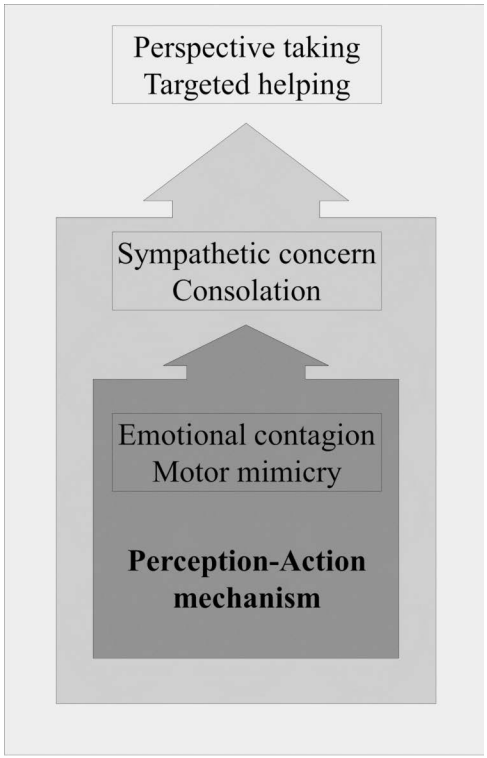


Figure 11.1 Russian doll model. Adapted from: de Waal and Preston (2017)

and targeted helping. From a developmental view, this model might look appealing at first sight, because it captures important aspects of the typical line of development of empathic abilities in human children, and we appreciate the ontogenetic perspective. Still, this conceptual framework is inadequate since the doll model implies (i) that the relevant cognitive abilities are unfolded one on top of the other and (ii) that they contribute to a full-blown case of empathy in the same way. (Ad i): Psychopathy is a counterexample to the presupposition of a sequential unfolding of these abilities: First-grade psychopaths typically lack sympathetic concern, but they are experts in perspective taking and, in principle, also in targeted helping although they might misuse this ability for achieving their own goals (Anderson and Kiehl 2014). (Ad ii): Furthermore, these components of the model do not belong to the same functional level: Sympathetic concern is an *attitude with a content* that is activated in typical cases of empathy, while perspective taking is a general *cognitive ability* that normally supports targeted helping. The latter enables an advanced self-other distinction, which is helpful both for the registration of the mental state of the observed person and for the response behaviour directed at this person. We are better off with a framework that allows for relevant cognitive abilities that are not necessarily grounded on each other, and we should aim for a characterisation that distinguishes different functional roles for cases of empathy.

A different approach is the combination model (Yamamoto 2017). He emphasises three aspects as typical *cognitive or behavioural goals* of empathy: matching with others, understanding others, and prosociality. The abilities often discussed in the context of empathy can be mapped

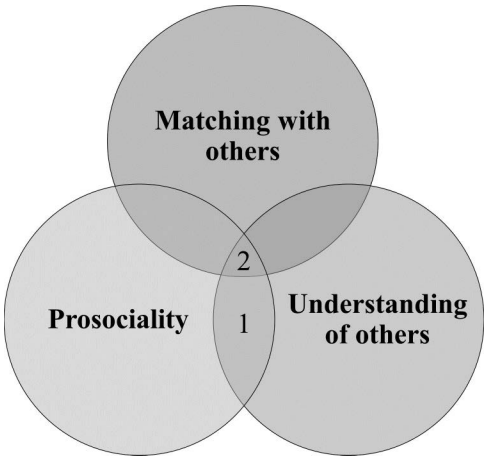


Figure 11.2 Combination model. Adapted from: Yamamoto (2017)

onto these three factors. Matching with others includes synchrony, mimicry, and emotional contagion. Understanding of others involves perspective taking, Machiavellian intelligence, and also Schadenfreude. Prosociality can be realised by food sharing or prosocial choice. In this model, more complex cases of empathy can be assigned to specific overlaps of factors, e.g. targeted helping consists of understanding others and prosociality (intersection 1). Finally, for sympathy, consolation, calculated reciprocity, and advantageous inequity aversion, all three factors are necessary (intersection 2). This approach avoids the problems of the Russian doll model since the three factors need not be sequentially based on each other and characterise different functional roles for sub-features involved in empathy. However, the model includes the claim that cases realizing only one factor can be categorised as clear cases of empathy even if these are basic ones. This is not convincing as can be illustrated by two examples: If prosocial behaviour is realised as an inborn reaction pattern or—more importantly—it is just understood as a certain type of beneficial behaviour independent from the *registration* of the mental state of the other and independent from the typical caring *attitude* toward the other, then this is not a case of empathy: The latter is, for example, realised if Tom is accidentally cooking the old lady's favourite meal just because he aims at becoming a specialised cook. He does not care for the old lady and forgot about her sadness. But after his cooking event, he brings a portion to every neighbour waiting for the evaluation of his cooking abilities and this involves bringing the old lady her favourite meal just at her birthday. This case is intuitively not a case of empathy. In our framework, this is accounted for because the behaviour is neither a response to registering the old lady's sadness about her vanishing cooking abilities nor acting with an attitude of caring for her. Only one of our three criteria is satisfied, namely a normally adequate prosocial behaviour. In our framework, it would be at best a borderline case but since the de facto prosocial behaviour is not even intended as such this shifts the case even more into a case of non-empathy. This example cannot be handled within the combination model conceptual framework because the relevant factors alone are not sufficient and prosociality as well as matching with others only describes the *behavioural response* in typical cases of empathy. Thus, this approach is too inclusive. We require the registration of the other's mental state as well as an attitude of caring for the other as relevant factors in some way. A second example that should be excluded from cases of empathy is motor mimicry, despite this being a typical case of matching with others: In these cases, the motor behaviour of another agent is copied, probably based on the activation of mirror neurons. These are specific neurons found in mammals, which are active not only when an action is performed by the subjects, but also when the subject perceives another individual performing the action (Iacoboni 2009). Evidence for this system outside the

mammalian lineage is scarce, but a study with songbirds suggests that a similar one might exist in birds and possibly other, older species (Miller 2008). This is an interesting behaviour but it should not be included as a case of empathy since it is only a matching motor reaction: Just matching with others' behaviour is not sufficient for empathy. This is adequately described in our framework because motor mimicry does not involve the registration of the other's affective state (or any mental state), nor any attitude directed toward the other. Furthermore, the behavioural response does not involve any relevant self-other distinction.

Thus, both kinds of approaches seem promising at first sight but open up to several problems. Hence, this chapter aims at finding a better conceptual framework of empathy.

A multi-component model of empathy

To develop a systematic overview of different forms of empathy, we need to clarify which kinds of phenomena we should account for. One central source of inspiration for our approach comes from developmental psychology, especially the ontogenetic model of Hoffman (2000). He defines five levels of empathy in child development: (1) Emotional Contagion, (2) Egocentric Empathic Distress, (3) Quasi-Egocentric Empathic Distress, (4) Veridical Empathic Distress, and (5) Empathic Distress Beyond the Concrete Situation. Number (4) in his model is what we call full-blown empathy and number (3) we call intermediate *empathy*. Inspired by this model, we aim at developing an analogous five-type model with a special focus on its applicability for animal cognition since we want to explicitly integrate the phylogenetic perspective into our theoretical framework and outline its fruitful application. But our multi-component model is different from his model since, first, it systematically distinguishes the core functional roles of registration, attitude, and behavioural response. Second, we characterise for each component the typical sub-criteria indicating possible implementations. Third, this enables us to precisely characterise the similarities and differences of the relevant phenomena and types of empathy.

The different forms of empathy are listed in the columns, rows are dedicated to underlying sub-criteria, whereby marking with 'x' means that a sub-criterion is involved in the phenomenon while an empty box expresses a non-involvement. Marking with '–' means that a criterion is not applicable to this form of empathy at all. In the case of pure cognitive empathy, the attributes 'positive' (pos.) and 'negative' (neg.) mark two different sub-phenomena of pure cognitive empathy.

We suggest the following three components and sub-criteria as motivated above: We can *register* the affective state, the relevant situation, and the mindset of the other. We can activate *the attitude of caring* for the other in a positive or negative way; and we realise a *behavioural*

Table 11.1 The multi-component model of empathy

		Motor mimicry	*Emotional contagion*	*Egocentric empathy*	*Intermediate empathy*	*Full-blown empathy*	*Future-oriented empathy*	*Pure cognitive empathy*	
								Pos	Neg.
Registration	Affective state	X	X	X	X	X	X-future	X	X
	(ii) Situation			X	X	X	X-future	X	X
	(iii) Mindset					X	X-future	X	X
Attitude of care for other					X	X	X	X	
Response	(a) Triggered affective state		X	X	X	X	X		
	(b) Relevant self-other distinction			X	X	X	X	X	X
	(c) Self-directed	—	—	X					
	(d) Other-directed	—	—		X	X	X	X	X

response, which typically is other-directed but can also be self-directed or may not even involve a relevant self-other distinction.

Most empathy researchers including Hoffman (2000) suggest that *emotional contagion (ideal-type 1)*, i.e. laughing when someone else is laughing, should be evaluated as a case of empathy. In our model, it is at best a borderline case since it fulfills only one out of three criteria. It involves the registration of the affective states of the other (in contrast to motor mimicry), followed by a transfer of this state onto the observer, thereby leading to an affective response matching that of the other. But this response is a fully automatic reaction to registering the affective state, thereby lacking the other two components, namely an attitude of caring and a relevant self-other distinction as being involved in the response behaviour. Emotional contagion is shown by new-born infants (Martin and Clark 1982) and is demonstrated in quite a variety of species, e.g. rodents (Langford et al. 2006), dogs (Yong and Ruffman 2014), and kea parrots (Schwing et al. 2017). From an evolutionary perspective, it is more demanding than just motor mimicry but it may also be mainly realised on the basis of mirror neurons since there is evidence for mirror neurons that are active in the case of expressions of pain and disgust (e.g. Wicker et al. 2003) and it is argued that there exists a mirror neuron system for emotions in general (Bastiaansen et al. 2009).

Ideal-type 2, which is the first clear but still basic form of empathic phenomena, is *egocentric empathy*. Typical phenomena include aversive behaviour or what may be called an 'as-if-it-was-me behaviour' (Hoffman 2000). From a developmental perspective, one concrete example observed in toddlers could be described as follows: One child falls and hits her knee. She is observed by another child, while she holds her knee and starts crying. The observing child now also holds his own knee and starts crying, as if he himself had been injured. A more basic behavioural response can be observed in the following case: One child A registers that another child B is very sad. But instead of helping B, A only seeks shelter with her mother to move away from the stressful challenge. This behaviour is shown by children of about ten months of age (Hoffman 1975; Zahn-Waxler et al. 1984). There is also evidence for a similar pattern in apes (de Waal 2009) and rodents (Meyza et al. 2017), demonstrating this kind of aversive behaviours toward distressed conspecifics. In our model, egocentric empathy involves two of three components, namely first, the registration of the affective state of the other and of the relevant situation; and second, the behavioural response. It involves a relevant self-other distinction, but remains self-directed, i.e. the agent tries to deal with the affective challenge by moving away from it (changing it for oneself) but not by changing it for the other. What is lacking is the attitude of caring for the other. One central feature to advance from pure emotional contagion to egocentric empathy is some clear form of registering the self-other distinction as it is e.g. realised in

the ability of joint attention developing in children between the age of nine and 12 months (Corkum and Moore 1998; Tomasello 1999). The recognition of the triadic relationship between self, other, and environment has also been shown for apes (Itakura 1996) using an eye-tracking paradigm. Furthermore, dogs are able to not only follow gazes, but also react differently depending on the other's attention (Marshall-Pescini et al. 2013), thereby suggesting the ability to recognise and maintain the triadic relationship. Taken together, this makes it plausible that the aversive behaviours cited above are cases of egocentric empathy.

Ideal-type 3, which we call *intermediate empathy*, involves direct helping behaviour as a typical phenomenon. In human ontogenesis, this can be observed in 12-month-old children's consolation behaviour (Zahn-Waxler et al. 1984). Another example is the following response to a crying friend: The boy first looked at the friend, then took her hand, but brought her to his own mother for comfort, although the friend's mother was also present (Hoffman 1978). A further step is made, if a child's observation of another person being sad, because she cannot reach her favourite toy, leads to direct helping, roughly starting when children are two years of age. In our model, this now involves all three components, namely the registration of the affective state and the relevant situation (sadness because of not getting the toy), and there is an attitude of caring for the other involved, as well as a response-behaviour that is other-directed (supporting him to get the toy) (Warneken 2006). Can we observe these behaviours in animals too? Chimps help a human experimenter to obtain an out-of-reach object, irrespective of whether or not they are offered a reward for doing so (Warneken et al. 2007). Capuchin monkeys do so in low-cost situations only (Barnes et al. 2008). Further evidence for this third type of empathy in animals comes from studies in which one animal observes a conspecific in distress and has the option to help by opening a cage or reducing negative effects. This helping behaviour is shown at least for primates (Yamamoto et al. 2012) and rodents (Sato et al. 2015). The animals demonstrate a motivation to help others instead of receiving a reward. We interpret this behaviour as an indicator that the attitude of caring is implemented since in the case of nonhuman animals we can only rely on behavioural indicators of an attitude.

Another interesting behaviour is consolation behaviour, which is less clear in its evaluation. Since it directly reduces not only the other's but also the observer's distress, it is not settled whether it is indeed an other-oriented response rather than a self-oriented one. There is evidence for consolation behaviour in apes (Romero et al. 2010; Clay and de Waal 2013), other primates (Palagi et al. 2014), as well as in dogs (Cools et al. 2008), wolves (Zimen 2003), elephants (Plotnik and de Waal 2014), and corvids (Seed et al. 2007), which would have to be examined in more detail to prove that it is other-oriented in nature across species.

Interestingly, as in humans, mostly the losers of a conflict are soothed in this way, supporting the hypothesis of empathic consolation as a case of intermediate empathy rather than gathering rank-specific advantages or own comfort only. At least for chimps, the behavioural biases for consolation are similar to those shown for human empathy, such as sex difference, or social closeness (Romero et al. 2010). This makes it likely that similar mechanisms are at work here, thereby supporting the hypothesis of consolation being an indicator for other-oriented empathic abilities.

Ideal-type 4 is full-blown empathy, which is typical for our everyday life. The important difference to type 3 of empathy is that the registration involves not only the affective state of the other and the relevant situation but in addition the registration of the other's mindset (which usually is different from one's own). Thus, it presupposes the ability of *cognitive perspective taking*, which is usually tested by the false-belief task, in addition to visual perspective taking. Let us think back to our example of Tom and the old lady: Even with the attitude to care for the other, the spontaneous response behaviour can only result in cooking the old lady's favourite meal as caring for her if you are aware of her preferences and account for them. If this ability is not yet developed, children may bring their own favourite toy as a birthday present for the mother, ignoring the different preferences. To investigate the sensitivity for others' mental states, there is a battery of false-belief tasks (FBT) established in developmental psychology, namely three implicit FBT (violation of expectation, anticipatory looking, active helping paradigm) and the classical explicit language-based FBT (de Bruin and Newen 2012). While the expectation violation paradigm is already passed by 15-month-old children (Onishi and Baillargeon 2005), and there is a debate about earlier observations, e.g. Kovacs et al. (2010), the active helping paradigm as an enactive form of implicit task is passed only at 18 months of age (Buttelmann et al. 2009); and the explicit FBT is not passed before age 4 (Wellman et al. 2001). To investigate evidence for sensitivity for the cognitive perspective of another agent in the realm of animals, we need to focus on nonlinguistic tests, i.e. the implicit FBTs. Is there evidence of full-blown empathy in animals? Yes, there is some direct and some indirect evidence, which taken together indicate that we can find full-blown empathy in some species. Krupenye et al. (2016) demonstrated that apes are able to pass the implicit FBT in the variant called the anticipatory looking paradigm. Although it is debated whether anticipatory looking demonstrates sensitivity for false belief, it is clear that it involves some sensitivity for false information of the other agent and this is sufficient for our argument. There is further evidence that this is not an over-interpretation since apes also pass the more demanding implicit FBT, namely the active helping FBT (Buttelmann et al. 2009, 2017), i.e. this helping behaviour presupposes that the apes are sensitive to the conspecific's being misinformed. Apes pass the active helping paradigm (Buttelmann et al. 2017), as they reliably open

only those boxes that the observed individual desires without specific perceivable cues, leading to the same result. In more detail, there is a recent analysis of passing the helping behaviour paradigm which does not presuppose full-blown belief representations: It can be modelled by ascribing the ability to represent not only regular mental files of a relevant object (where it is located or what preference one has for it), but also perspectival mental files of the same object (where the other thinks it is or what preference the other has for it) (Newen and Wolf 2020). At least great apes (bonobo, chimpanzee, and orangutan) have been shown to pass two demanding implicit FBTs: Thus, the first component of registering the cognitive state of others is realised, at least to an interesting degree. Furthermore, the active helping paradigm indicates the attitude of caring for the other, but we have to admit that having evidence for the attitude independent from the response behaviour is an open question when applying the framework to animals. The third component is a response behaviour that accounts for the others' affective state. The helping activities clearly account for the desire expressed by the other to get a certain object. Thus, this indicates that at least great apes are able to act with full-blown empathy. For other species this mainly remains to be investigated systematically. Furthermore, this evaluation needs addtional tests or even may be revised due to the ongoing replication crisis of the FBTs (Kulke et al. 2018).

Indirect evidence for an understanding of the other's mental states comes from research on the awareness of the momentary abilities of the other in tasks distinguishing unwillingness and inability to fulfill the task. At least primates and birds understand the difference whether a specific individual is unable or unwilling to perform certain actions (capuchin monkeys: Phillips et al. 2009; macaques: Canteloup and Meunier 2017; New Caledonian crows: Taylor et al. 2012), thereby showing the ability to understand the cognitive potential of the other. Further evidence comes from ravens that recognise the others' visual access solely on the basis of auditory input (Bugnyar et al. 2016). Taken together, there is quite some evidence that at least the helping behaviour of chimpanzees could be interpreted as a case of full-blown empathy. For other species we can observe some evidence for sensitivity for the cognitive perspective of others, but this is in need of further investigation before one can settle the question whether more species are able to realise full-blown empathy.

Ideal-type 5 of empathic abilities is what we call *future-oriented empathy*. Here, the affective information is typically overridden by information directed at the near or far future of the other, thereby leading to a response not trying to improve the other's current situation, but her future; e.g. if a child is sad because the mother insists that she does her homework, the mother may continue to insist despite noticing the sadness of her daughter because she anticipates the relevance of success in school for her in a few years. The mother anticipates that she will be happy about the success later and therefore does not do something to

stop the sadness now. Hence, the capacity for future planning and an adaptation to the other is necessary.

The additional ability needed here is episodic foresight. Evidence from developmental psychology shows a strong connection of episodic foresight and episodic memory. Children are able to think about others' future and their own already at the age of three years (Payne et al. 2015) in a basic form; it unfolds into more sophisticated forms until reaching an adult level at the age of 14–16, strongly depending on the personal background and other perspective-taking abilities (Hoffman 2000). Since both abilities seem to be so strongly interconnected, and even rely on overlapping brain circuits (Addis et al. 2007; Szpunar et al. 2007; for a critical review see Schacter et al. 2012), evidence for episodic memory can be taken as an indication of episodic foresight. Can we observe episodic foresight in animals? Raby et al. (2007) confirmed episodic memory in western scrub jays by using a what-where-when paradigm. The birds were not only able to remember what they cached where and when, but they could also operate flexibly on this information, thereby indicating a capacity for episodic-like memory. This result was further investigated and confirmed by Correia et al. (2007), where scrub jays could not only use their memory to plan for the future, but also did so if their current motivational state would suggest a different behaviour, meaning that the birds can act according to their future needs even against their current desires, thereby providing good evidence to reject the Bischof-Köhler-Hypothesis. This goes even beyond the usage of episodic memory in the here and now, since in the mentioned study it is necessary to combine episodic memory with the anticipation of future needs to plan for the future. The ability to form and use what-where-when memory was also tested and confirmed in other corvid species such as magpies (Zinkivskay et al. 2009), as well as in rodents (Babb and Crystal 2005; Crystal 2013). Interestingly, evidence for the what-where-when memory in chimpanzees is still missing (Dekleva et al. 2011). This evidence for episodic foresight only hints at future-oriented empathy if it can be combined with strong indications for cognitive perspective taking. This was tested by Stulp et al. (2009), showing that scrub jays react across different modalities when being observed during caching. This implies that there is an amodal representation of being observed, which illustrates an amodal version of perspective taking. Additionally, Ostojic et al. (2013) found that male jays can recognise desire-states of female jays during food sharing based on registering what the female had been fed last; their food offer accounts for the female's preferences (at least partially), inhibiting their own preference. Taken together, this evidence supports the hypothesis that animals are able to anticipate future states of others and act flexibly on this information.

There are also some single case studies describing how apes use future-oriented empathy to prevent harm to others before the harmful situation even occurs (de Waal 2009; Hirata 2009).

Even though some nonhuman animals possess the capacity for primitive forms of future-oriented empathy, the more sophisticated ones appear to be uniquely human. A typical example is the situation of an individual observing a disabled child playing happily in the sandbox. The observer might commiserate even though the child is in a happy state at the moment. This is because of the knowledge of the future of the child, how it will be acted upon, and on which things of life it might miss out. Here, it is necessary to perceive others as persons with personal pasts and futures extending over several years, which seems to be an ability uniquely human.

Special cases of empathy and related phenomena

There are several *special cases* of empathy, which can also be included within the proposed framework. One of these that we want to discuss in more detail is *pure cognitive empathy*. One characteristic feature is that it involves the registration of the other's affective state but in this case it does not lead to a corresponding affective state in the observer or at least the corresponding response behaviour is not based on an affective state in the observer. We can distinguish at least a positive and a negative case of pure cognitive empathy in the sense that care for others is involved in the realisation or not.

Let us start with *the positive case*: Paradigmatic everyday cases of pure cognitive empathy are given when professionals try to help their clients who are in bad physical or psychological condition. In such cases, surgeons or, respectively, psychiatrists need to register the physical and/or psychological situation, including the affective state of the client; they could do this by relying on a cognitive evaluation of the situation or try to downgrade or even inhibit their own affective involvement to keep their professional distance. Despite having an adequate understanding of the affective situation of the other and maybe also expressing this, there needs to be implemented a sense of distancing with an appropriate level of detachment from the others' emotion (Hojat et al. 2002). This may still be accompanied with an attitude of caring for the other and with a supporting response behaviour, but if there is no corresponding affective state in the observer, then this is either just based on a pure rational evaluation or it is based on a strong inhibition of a disposition for an affective response. Thus, positive cognitive empathy is paradigmatically realised in such a professional stance which seems to be often involved in the treatment of patients in the long run. In comparison to full-blown empathy in the positive versions of cognitive empathy, only one feature is missing, namely the affective involvement of the observer.

A negative case of pure cognitive empathy is realised in some version of psychopathy. Psychopathy is a complex mental disorder with a rich cluster of symptoms. Already through childhood and youth, lack

of emotionality together with high aggression potential and intentional instrumentalizing violence is characteristic for this disorder (Frick and Marsee 2006; Vitacco and Vincent 2006). We focus just on primary psychopaths, also called 'low-anxious psychopaths'. We classify their typical behaviour as a case of negative cognitive empathy, because in comparison to positive cognitive empathy the attitude of caring for others is missing. This leads to the well-known characteristic that psychopaths can easily use their knowledge about others to take actions of utilizing others irrespective of the potential harm and distress caused without remorse or intrinsic inhibition. The relevant underlying features we want to highlight are the following: (i) Psychopaths experience their own negative emotions and especially fear in an extremely reduced way (Del Gaizo and Falkenbach 2008, 209). They do not get 'infected' by the stress sweat of other people; non-psychopaths, in contrast, get anxious when they breathe in others' stress sweat (Dutton 2013, 20, 45ff). (ii) Nevertheless, psychopaths are often as good as non-psychopaths in identifying the emotions of others: On the 'Reading the mind in eyes' test, psychopaths and people with psychopathic traits are as good as non-psychopaths in identifying the emotions of other people. On this test, the subjects have to identify emotions by means of photographs of the eye area of others. People with autism, by comparison, make a lot of mistakes on the 'Reading the mind in eyes' test (Richell et al. 2003, 525).[4] Thus, a first-grade psychopath can adequately observe the other's affective state, situation, and mindset but he does not automatically activate a complementary affective state. Thus, he can easily choose an attitude of not caring for the other but only for himself and thus misuse the knowledge of the other for manipulations. For these special cases like cognitive empathy, one may also ask when they show up in ontogenetic development and whether animals could also have them. But here we constrain ourselves to outline the conceptual possibilities of special cases not yet mentioned in our list of types of empathy to indicate that there is room for many more combinations of sub-features. This involves the possibility to systematically describe many related phenomena often discussed in the context of empathy too. Adriaense et al. (2020) mention e.g. 'Knowing Another Person's Internal State, Including His or Her Thoughts and Feelings' (64): In our model, this can be equated with the *registration dimension* of full-blown empathy. Thus, we offer a promising conceptual framework.

Summary and conclusion

Our multi-component account allows us to characterise a variety of ideal types of empathy from an evolutionary perspective: We exclude motor mimicry and characterise emotional contagion as a borderline case, while there are three typical cases of empathy: egocentric, intermediate, and full-blown empathy. These types of empathy are characterised

in detail by relying on our three components of registering, caring, and acting out a behavioural response. Since each component is analysed as based on typical sub-features, we are able to describe how the demands of more complex forms of empathy systematically increase before reaching full-blown empathy. But this account does not only allow us to tell a standard phylogenetic and ontogenetic accumulation story, but also, since typical features of full-blown empathy can be realised in a variety of ways or some can even be lacking, our account should be understood as describing profiles of empathy with the variants in Table 11.1 as ideal-typical steps in a normal development. But profiles can vary and thus we can expect to observe special cases of empathy like pure cognitive empathy. Using our multidimensional account of empathy for a comparative perspective we are able to claim that the first four types of phenomena can be found in animals, where the more demanding the type of empathy the fewer species can implement it. Full-blown empathy can be reasonably observed in great apes and may be expected in other social animals which are candidates for the ability of cognitive perspective taking. But this needs to be investigated in future research. We think that our conceptual framework is a preferable tool in contrast to existing theories of empathy that allows us to start systematic empirical investigation and comparisons of empathy in humans and nonhuman animals.

Acknowledgements

The authors acknowledge Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer GRK-2185/2 (DFG-Graduiertenkolleg Situated Cognition). Furthermore, this is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the project (NE 576/14-1) 'The structure and development of understanding actions and reasons'.

Notes

1 A short version with overlapping ideas and overlapping sections is published in the conference proceedings of the Cognitive Science Society 2021: Newen, A. and Griem, M. 2021. A Conceptual Framework for Empathy and Its Application to Investigate Nonhuman Animals, *Proceedings of the Cognitive Science Society*. The Sections 'Criteria of adequacy for a concept of empathy developed from a typical case' and 'Criticizing two central conceptual frameworks' are mainly overlapping, the other sections are different.

2 This is taken from: https://dictionary.cambridge.org/de/worterbuch/englisch/empathy, last accessed 30 March 2022.

3 This needs a short explication: To avoid over-intellectualizing this component, we allow for non-conceptual attitudes of caring (in line with Glock's characterisation of nonrepresentational types of thinking, Glock 2013); and we need to distinguish the case in which someone is not able to develop an

attitude of caring (cases of non-applicability of the caring attitude) from the case in which someone is able to care for the other but decides not to do so; the latter is realised in the case of psychopathy when the psychopath misuses an adequate understanding of the relevant affective state, situation, and mindset of the other to hinder and not to help.

4 This asymmetry in a high-level ability of observing the emotions of others and a low-level ability of experiencing emotions is a challenge for mirror neuron theorists. One suggested solution is that psychopaths normally have strongly reduced pain experience while witnessing the pains of others. When asked to empathise, however, they can activate their own feeling. They seem to be able to switch a corresponding affective state on which is normally not activated. Thus, mirror neurons may play a complex role but the observation is not a knock-down observation for mirror neuron theorists (Meffert et al. 2013).

References

Addis, D.R., Wong, A.T. and Schacter, D.L. (2007). Remembering the Past and Imagining the Future: Common and Distinct Neural Substrates during Event Construction and Elaboration. *Neuropsychologia* 45(7): 1363–1377. https://doi.org/10.1016/j.neuropsychologia.2006.10.016

Adriaense, J.E.C. et al. (2020). Challenges in the Comparative Study of Empathy and Related Phenomena in Animals. *Neuroscience & Biobehavioral Reviews* 112: 62–82. https://doi.org/10.1016/j.neubiorev.2020.01.021

Anderson, N.E. and Kiehl, K.A. (2014). Psychopathy: Developmental Perspectives and Their Implications for Treatment. *Restorative Neurology and Neuroscience* 32(1): 103–117. https://doi.org/10.3233/RNN-139001

Babb, S.J. and Crystal, J.D. (2005). Discrimination of What, When, and Where: Implications for Episodic-Like Memory in Rats. *Learning and Motivation* 36(2): 177–189. https://doi.org/10.1016/j.lmot.2005.02.009

Barnes, J.L. et al. (2008). Helping Behaviour and Regard for Others in Capuchin Monkeys (*Cebus apella*). *Biological Letters* 4: 638–640. https://doi.org/10.1098/rsbl.2008.0410

Bartal, I.B.-A., Decety, J. and Mason, P. (2011). Empathy and Pro-Social Behavior in Rats. *Science* 334(6061): 1427–1430. https://doi.org/10.1126/science.1210789

Bastiaansen, J.A.C.J., Thioux, M. and Keysers, C. (2009). Evidence for Mirror Systems in Emotions. *Philosophical Transactions of the Royal Society B Biological Sciences* 364: 2391–2404. https://doi.org/10.1098/rstb.2009.0058

Bischof-Köhler, D. (1985). Zur Phylogenese menschlicher Motivation. In Eckensberger, L.H. and Baltes, M.M. (eds.), *Emotion und Reflexivität*. München: Urban & Schwarzenberg, pp. 3–47.

Bugnyar, T., Reber, S.A. and Buckner, C. (2016). Ravens Attribute Visual Access to Unseen Competitors. *Nature Communications* 7: 10506. https://doi.org/10.1038/ncomms10506

Buttelmann, D., Carpenter, M., and Tomasello, M. (2009). Eighteen-Month-Old Infants Show False Belief Understanding in an Active Helping Paradigm. *Cognition* 112(2): 337–342. https://doi.org/10.1016/j.cognition.2009.05.006

Buttelmann, D. et al. (2017). Great Apes Distinguish True from False Beliefs in an Interactive Helping Task. *PLOS ONE* 12: e0173793. https://doi.org/10.1371/journal.pone.0173793

Canteloup, C. and Meunier, H. (2017). 'Unwilling' versus 'Unable': Tonkean Macaques' Understanding of Human Goal-Directed Actions. *PeerJ* 5: e3227. https://doi.org/10.7717/peerj.3227

Clay, Z. and de Waal, F.B.M. (2013). Bonobos Respond to Distress in Others: Consolation across the Age Spectrum. *PLOS ONE* 8: 55206. https://doi.org/10.1371/journal.pone.0055206

Cools, A.K.A., van Hout, A.J.-M. and Nelissen, M.H.J. (2008). Canine Reconciliation and Third-Party-Initiated Postconflict Affiliation: Do Peacemaking Social Mechanisms in Dogs Rival Those of Higher Primates? *Ethology* 114(-1): 53–63. https://doi.org/10.1111/j.1439-0310.2007.01443.x

Corkum, V. and Moore, C. (1998). The Origins of Joint Visual Attention in Infants. *Developmental Psychology* 34(1): 28.

Correia, S.P.C., Dickinson, A., and Clayton, N.S. (2007). Western Scrub-Jays Anticipate Future Needs Independently of Their Current Motivational State. *Current Biology* 17(10): 856–861. https://doi.org/10.1016/j.cub.2007.03.063

Crystal, J.D. (2013). Remembering the Past and Planning for the Future in Rats. *Behavioural Processes* 93: 39–49. https://doi.org/10.1016/j.beproc.2012.11.014

de Bruin, L.C. and Newen, A. (2012). An Association Account of False Belief Understanding. *Cognition* 123(2): 240–259. https://doi.org/10.1016/j.cognition.2011.12.016

de Waal, F.B.M. (2009). *The Age of Empathy: Nature's Lessons for a Kinder Society*. London: Three Rivers Press.

de Waal, F.B.M. and Preston, S.D. (2017). Mammalian Empathy: Behavioural Manifestations and Neural Basis. *Nature Reviews Neuroscience* 18: 498–509. https://doi.org/10.1038/nrn.2017.72

Dekleva, M. et al. (2011). Chimpanzees (*Pan Troglodytes*) Fail a What-Where-When Task but Find Rewards by Using a Location-Based Association Strategy. *PLOS ONE* 6: 16593. https://doi.org/10.1371/journal.pone.0016593

Del Gaizo, A.L. and Falkenbach, D.M. (2008). Primary and Secondary Psychopathic-Traits and Their Relationship to Perception and Experience of Emotion. *Personality and Individual Differences* 45(3): 206–212. https://doi.org/10.1016/j.paid.2008.03.019

Dutton, K. (2013). *Psychopathen: Was man von Heiligen, Anwälten und Serienmördern lernen kann*. München: Deutscher Taschenbuch Verlag.

Feshbach, N.D. and Feshbach, S. (2009). Empathy and Education. In Decety, J. and Ickes, W. (eds.), *The Social Neuroscience of Empathy, Social Neuroscience*. Cambridge, MA: MIT Press, pp. 85–97. https://doi.org/10.7551/mitpress/9780262012973.003.0008

Frick, P.J. and Marsee, M.A. (2006). Psychopathy and Developmental Pathways to Antisocial Behavior in Youth. In Patrick, C.J. (ed.), *Handbook of Psychopathy*. New York, NY: The Guilford Press, pp. 353–374.

Glock, H.J. (2000). Animals, Thoughts and Concepts. *Synthese* 123: 35–64.

Glock, H.J. (2013). Animal Minds: A Non-Representationalist Approach. *American Philosophical Quarterly* 50(3): 213–232.

Glock, H.J. (2019). Agency, Intelligence and Reasons in Animals. *Philosophy* 94: 645–671. https://doi.org/10.1017/S0031819119000275

Hirata, S. (2009). Chimpanzee Social Intelligence: Selfishness, Altruism, and the Mother–Infant Bond. *Primates* 50: 3–11. https://doi.org/10.1007/s10329-008-0122-1

Hoffman, M.L. (1975). Moral Internalisation, Parental Power, and the Nature of Parent-Child Interaction. *Developmental Psychology* 11(2): 228–239. https://doi.org/10.1037/h0076463

Hoffman, M.L. (1978). Psychological and Biological Perspectives on Altruism. *International Journal of Behavioral Development* 1: 323–339. https://doi.org/10.1177/016502547800100403

Hoffman, M.L. (2000). *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511805851

Hojat, M. et al. (2002). Empathy in Medical Students as Related to Academic Performance, Clinical Competence and Gender. *Medical Education* 36(6): 522–527.

Iacoboni, M. (2009). Imitation, Empathy, and Mirror Neurons. *Annual Review of Psychology* 60: 653–670. https://doi.org/10.1146/annurev.psych.60.110707.163604

Itakura, S. (1996). An exploratory Study of Gaze-Monitoring in Nonhuman Primates. *Japanese Psychological Research* 38(3): 174–180. https://doi.org/10.1111/j.1468-5884.1996.tb00022.x

Kovacs, A.M., Teglas, E. and Endress, A.D. (2010). The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science* 330(6012): 1830–1834. https://doi.org/10.1126/science.1190792

Krupenye, C., et al. (2016). Great Apes Anticipate That Other Individuals Will Act According to False Beliefs. *Science* 354(6308): 110–114. https://doi.org/10.1126/science.aaf8110

Kulke, L., B. von Duhn, D. Schneider, and H. Rakoczy. 2018. Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. Psychological Science 29 (6): 888–900. https://doi.org/10.1177/0956797617747090.

Langford, D.J. et al. (2006). Social Modulation of Pain as Evidence for Empathy in Mice. *Science* 312(5782): 1967–1970. https://doi.org/10.1126/science.1128322

Laurence, S. and Margolis, E. (1999). Concepts and Cognitive Science. In Laurence, S. and Margolis, E. (eds.), *Concepts: Core Readings*. Cambridge, MA: MIT Press, pp. 3–81.

Marshall-Pescini, S. et al. (2013). Gaze Alternation in Dogs and Toddlers in an Unsolvable Task: Evidence of an Audience Effect. *Animal Cognition* 16: 933–943. https://doi.org/10.1007/s10071-013-0627-x

Martin, G.B. and Clark, R.D. (1982). Distress Crying in Neonates: Species and Peer Specificity. *Developmental Psychology* 18(1): 3–9. https://doi.org/10.1037/0012-1649.18.1.3

Meffert H. et al. (2013). Reduced Spontaneous But Relatively Normal Deliberate Vicarious Representations in Psychopathy. *Brain* 136: 2550–2562. https://doi.org/10.1093/brain/awt190.

Meyza, K.Z. et al. (2017). The Roots of Empathy: Through the Lens of Rodent Models. *Neuroscience and Biobehavioral Reviews* 76: 216–234. https://doi.org/10.1016/j.neubiorev.2016.10.028

Miller, G. (2008). Neuroscience. Mirror Neurons May Help Songbirds Stay in Tune. *Science* 319(5861): 269. https://doi.org/10.1126/science.319.5861.269a

Newen, A. and Bartels, A. (2007). Animal Minds and the Possession of Concepts. *Philosophical Psychology* 20(3): 283–308.

Newen, A. and Wolf, J. (2020). The Situational Mental File Account of the False Belief Tasks: A New Solution of the Paradox of False Belief Understanding. *Review of Philosophy and Psychology* 11: 717–744. https://doi.org/10.1007/s13164-020-00466-w

Onishi, K.H. and Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science* 308(5719): 255–258. https://doi.org/10.1126/science.1107621

Ostojic, L., et al. (2013). Evidence Suggesting That Desire-State Attribution May Govern Food Sharing in Eurasian Jays. *Proceedings of the National Academy of Sciences of the United States of America* 110(10): 4123–4128. https://doi.org/10.1073/pnas.1209926110

Palagi, E. et al. (2014). Exploring the Evolutionary Foundations of Empathy: Consolation in Monkeys. *Evolution and Human Behavior* 35(4): 341–349. https://doi.org/10.1016/j.evolhumbehav.2014.04.002

Payne, G., et al. (2015). Mental Time Travel for Self and Other in Three- and Four-Year-Old Children. *Memory* 23(5): 675–682. https://doi.org/10.1080/09658211.2014.921310

Phillips, W. et al. (2009). 'Unwilling' versus 'Unable': Capuchin Monkeys' (*Cebus apella*) Understanding of Human Intentional Action. *Developmental Science* 12(6): 938–945. https://doi.org/10.1111/j.1467-7687.2009.00840.x

Plotnik, J.M. and de Waal, F.B.M. (2014). Asian Elephants (*Elephas maximus*) Reassure Others in Distress. *PeerJ* 2, e278. https://doi.org/10.7717/peerj.278

Raby, C.R. et al. (2007). Planning for the Future by Western Scrub-Jays. *Nature* 445: 919–921. https://doi.org/10.1038/nature05575

Richell, R.A. et al. (2003). Theory of Mind and Psychopathy: Can Psychopathic Individuals Read the 'Language of the Eyes'? *Neuropsychologia* 41(5): 523–526. https://doi.org/10.1016/s0028-3932(02)00175-6

Romero, T., Castellanos, M.A., and de Waal, F.B.M. (2010). Consolation as Possible Expression of Sympathetic Concern among Chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America* 107(27): 12110–12115. https://doi.org/10.1073/pnas.1006991107

Sato, N. et al. (2015). Rats Demonstrate Helping Behavior Toward a Soaked Conspecific. *Animal Cognition* 18: 1039–1047. https://doi.org/10.1007/s10071-015-0872-2

Schacter, D.L. et al. (2012). The Future of Memory: Remembering, Imagining, and the Brain. *Neuron* 76: 677–694. https://doi.org/10.1016/j.neuron.2012.11.001

Schwing, R. et al. (2017). Positive Emotional Contagion in a New Zealand Parrot. *Current Biology* 27(6): R213–R214. https://doi.org/10.1016/j.cub.2017.02.020

Seed, A.M., Clayton, N.S. and Emery, N.J. (2007). Postconflict Third-Party Affiliation in Rooks, *Corvus frugilegus*. *Current Biology* 17(2): 152–158. https://doi.org/10.1016/j.cub.2006.11.025

Stulp, G. et al. (2009). Western Scrub-Jays Conceal Auditory Information When Competitors Can Hear but Cannot See. *Biology Letters* 5(5): 583–585. https://doi.org/10.1098/rsbl.2009.0330

Szpunar, K.K., Watson, J.M. and McDermott, K.B. (2007). Neural Substrates of Envisioning the Future. *Proceedings of the National Academy of Sciences of the United States of America* 104: 642–647. https://doi.org/10.1073/pnas.0610082104

Taylor, A.H., Miller, R. and Gray, R.D. (2012). New Caledonian Crows Reason about Hidden Causal Agents. *Proceedings of the National Academy of Sciences* 109: 16389–16391. https://doi.org/10.1073/pnas.1208724109

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Vitacco, M.J. and Vincent, G.M. (2006). Understanding the Downward Extension of Psychopathy to Youth: Implications for Risk Assessment and Juvenile Justice. *International Journal of Forensic Mental Health* 5(1): 29–38. https://doi.org/10.1080/14999013.2006.10471228

Warneken, F. (2006). Altruistic Helping in Human Infants and Young Chimpanzees. *Science* 311(5765): 1301–1303. https://doi.org/10.1126/science.1121448

Warneken, F. et al. (2007). Spontaneous Altruism by Chimpanzees and Young Children. *PLoS Biology* 5(7): e184. https://doi.org/10.1371/journal.pbio.0050184

Wellman, H.M., Cross, D. and Watson, J. (2001), Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development* 72: 655–684. https://doi.org/10.1111/1467-8624.00304

Wicker, B., et al. (2003). Both of Us Disgusted in *My* Insula. *Neuron* 40(3): 655–664. https://doi.org/10.1016/S0896-6273(03)00679-2

Yamamoto, S. (2017). Primate Empathy: Three Factors and Their Combinations for Empathy-Related Phenomena: Primate Empathy. *WIREs Cognitive Science* 8(3): e1431. https://doi.org/10.1002/wcs.1431

Yamamoto, S., Humle, T. and Tanaka, M. (2012). Chimpanzees' Flexible Targeted Helping Based on an Understanding of Conspecifics' Goals. *Proceedings of the National Academy of Sciences of the United States of America* 109(9): 3588–3592. https://doi.org/10.1073/pnas.1108517109

Yong, M.H. and Ruffman, T. (2014). Emotional Contagion: Dogs and Humans Show a Similar Physiological Response to Human Infant Crying. *Behavioural Processes* 108: 155–165. https://doi.org/10.1016/j.beproc.2014.10.006

Zahn-Waxler, C. et al. (1984). Altruism, Aggression, and Social Interactions in Young Children with a Manic-Depressive Parent. *Child Development* 55: 112–122. https://doi.org/10.2307/1129838

Zimen, E. (2003). *Der Wolf: Verhalten, Ökologie und Mythos; das Vermächtnis des bekannten Wolfsforschers*. Stuttgart: Kosmos.

Zinkivskay, A., Nazir, F. and Smulders, T.V. (2009). What-Where-When Memory in Magpies (*Pica pica*). *Animal Cognition* 12: 119–125. https://doi.org/10.1007/s10071-008-0176-x

12 Plants, wants, and agents

Helen Steward

Hans-Johann Glock has long been a dauntless champion of the cognitive and other psychological capacities of animals.[1] His determination to combat the implausible exaggerations typical of some linguistically-oriented philosophy has been all the more powerful as a result of the fact that he himself has done much work within that linguistically-focused tradition, and so is in a position to appreciate its considerable strengths and subtleties, as well as the limitations to which he draws attention. It is precisely because he understands so well what there is to be said in favour of the position he calls 'lingualism', which claims that animals are confined to 'mere behaviour' and cannot engage in truly intentional action, that he has been able to argue so effectively against it, defending in his most recent work on the subject (Glock 2019) the still controversial view that a great many non-human animals—those Glock calls the 'higher' animals[2]—can not only act, but also act intelligently, rationally, and for reasons.

I intend simply to accept, for the purposes of this chapter, that Glock is correct that there are many kinds of animal that pass the test for these more sophisticated kinds of agency with flying colours—and let us for convenience's sake agree with Glock also in calling those that do the 'higher' animals, even if the terminology owes more than is really desirable to discredited ideas about the 'order' of nature. In this chapter, though, I want to focus not on these impressive creatures—the chimpanzees, the dolphins, the elephants, and the like—but rather on the vast range of cognitively less sophisticated organisms about which our intuitive verdicts concerning the presence even of basic agency might be less clear, or even determinately negative—including a wide range of 'lower' animals; the entirety of the plant kingdom; and an assortment of other living entities, such as bacteria, viruses, and fungi. In his (2019), Glock offers some brief suggestions as to how we might make what he claims are two important distinctions amongst these various kinds of entity: (i) between those capable of what Glock calls 'behaviour', on the one hand, and those that are merely 'processors of information', on the other; and (ii) between the *mere* behavers, on the one hand, and those capable of something we might want rather to dignify with the name

DOI: 10.4324/9781003202929-16

of *agency*, on the other. My aim here will be to describe and criticise in certain respects Glock's views concerning these interesting distinctions.

Before moving to do so, however, let me explain why I take Glock's work to be distinctive and important, by setting his work on animal agency in the context of a spectrum of alternative views. Glock has mainly tended to concentrate his efforts in this part of his philosophical endeavour on opposing what one might call 'inflationary' conceptions of agency. One kind of inflationary conception defines the power of agency itself in such a way as to make it coincide with a concept-drenched, thought-infused, way of impinging on the world, so that nothing that cannot be a wielder of concepts can be an agent either. On this inflationary view, to which language-possession is central, agency is plausibly already a capacity beyond the reach of most non-human animals, perhaps even beyond the reach of any.[3] Another variety of inflationary conception, which also comes within Glock's sights, identifies agency *per se* with fully rational, or at least intentional agency. There are different views about what an intentional agent is—but some such views would likely put agency beyond the reach of many simpler animals such as spiders, requiring, for example, that an intentional agent must at the very least possess a range of particular kinds of mental state, such as beliefs, desires, and intentions; states which one might not want to ascribe to these simpler animals, for a variety of possible reasons.[4] Glock has rightly protested against the excessive 'differentialism' (2010, 384) as he calls it, implied by positions of this sort. In recent years, however, it is notable that alternative *deflationary* conceptions of agency have begun to gain ground, occupying the opposite end from the inflationist position of the spectrum of views about what agency requires. Deflationary conceptions, often frequently motivated by environmentalist thinking, as well as by improvements in our understanding of the surprising behavioural complexity and flexibility of many kinds of 'lower' organism, offer, by contrast to inflationism, a thoroughly *gradualist* approach to agential power. A now burgeoning literature argues for the presence of at least some version of agency (as well as cognition, perception, and a range of other mental capacities and states) across most, if not the whole, of the living world.[5] Representations have been made on behalf of many kinds of insect (Barron and Klein 2016; Chittka 2017; Howard et al. 2018; Chittka and Wilson 2019; Ginsburg and Jablonka 2019) and there are also defenders of the idea that something worth regarding as a minimal version of agency is also to be found in plants (Hall 2011; Gagliano 2014; Mancuso and Viola 2015; Mancuso 2017); and even in single-celled organisms (Marshall 2009; Bassler 2018). The non-living world is often also incorporated into the deflationist's picture—talk of 'artificial intelligence' and 'artificial agents' is already widespread and there are many who see no reason to suppose that such talk is in any way metaphorical or in need of qualification. On deflationary views, agency is not a distinctive power, unique to a certain subset of animals, but rather

something which characterises life (including sometimes even artificial life) itself, though admittedly with differing degrees of sophistication. We are thereby relieved in one fell swoop of the need to decide (even roughly) where the mere behavers end and the agents begin, there being no such place in nature; and we become more appropriately attuned, it might seem, to the continuities characteristic of the living world.

My interest in Glock's work stems from the fact that though he is an anti-inflationist, he is clearly no deflationist either. He is attempting to occupy what I regard as promising, but easily destabilised middle ground between inflationism and deflationism, endorsing the idea that agency is not, in fact, the common property of all life, but rather a distinctive evolutionary development, a power that characterises some organisms (including many non-human ones) but not others. The big question, as I see it, is whether this kind of intermediate position can be successfully defended against the universalist gradualism of the deflationists. My suggestion is that though it may perhaps turn out not to be so, it would be premature to give up on the rather natural idea that the answer is 'yes'. Glock offers what I think is a rare vision of how precisely we might prosecute the goal of developing a conception of agency as a *non-universal* but also *non-human-specific* property of biological (and perhaps other information-processing) systems—and although I shall be arguing that the vision is wanting in certain respects, the general goal of trying to occupy this tricky area of logical space is one I share.

The chapter falls into two halves. In the first, I shall focus on Glock's distinction between mere information processing and behaviour, registering some doubts about whether the distinction is really properly objective, in a sense I shall attempt to explain. In the second half, I shall turn to look at the distinction that Glock makes between mere behaviour and agency, focusing in particular on the idea, developed in his (2019), that some animals have *wants* which may conflict with their biological needs, and Glock's suggestion that it is this property which marks out the agents from the rest. I shall suggest that in choosing this criterion as the mark of a real agent, Glock has, in fact, repeated one of the mistakes of which he finds inflationists guilty, in failing to see that there is such a thing as *basic* agency, which falls below the threshold indicated by the 'want' criterion, but yet deserves recognition as going beyond mere behaviour in some very important ways. I shall then conclude with the suggestion that a distinction which Glock considers *en passant*, only to set it aside, looks much closer to being what we might need, in order to help characterise the basic agents, than the capacity for wants.

From information processing to behaviour

Glock embarks on his discussion of the concept of behaviour from the traditional Aristotelian distinction between plants and animals. On the

Aristotelian view, plants have the capacities for nutrition and growth and fulfil their needs for light, water, and the very important requirement not to be prematurely (or too comprehensively) eaten in various ways, by growing roots, orienting their leaves in certain directions, growing thorns, and secreting stinging chemicals to deter herbivores, etc. Animals also have the capacities for nutrition and growth—but in addition possess two further notable classes of power—which Glock characterises as 'perception' and 'locomotion'. Glock claims that there is no *conceptual* connection between these two capacities, on the one hand, and the biological conception of animality, on the other (which he appears to regard as based on the distinction between autotrophs and heterotrophs)—there is nothing contradictory, he points out, about the idea of an autotrophic organism that nevertheless perceives and moves; or of a heterotrophic organism that is nevertheless rooted to the spot and lacking in any capacity for perception. Nevertheless, he writes:

> [. . .] the Aristotelian taxonomy pinpoints a distinctive type of activity, which, for biological reasons, is the rule among animals, and the exception among plants. It is activity in pursuit of needs through movement guided by perception. The terminus of such activity is its *goal. And although biologists are happy to speak of plant 'behaviour', I shall reserve that label for animate activity that pursues a goal informed by perception*.
>
> (Glock 2019, 649, my italics)

It appears, then, that behaviour, for Glock, is definitionally confined to the animal kingdom (at least on the assumption that that is what the word 'animate' seeks to convey); mere plants cannot 'behave' in the sense of the word he aims to clarify.

Choosing to 'reserve the label' of 'behaviour' only for animals, of course, cannot (at any rate without more argument) be regarded as the statement of a philosophical position. In the absence of further support, the claim that plants (for example) do not 'behave' would be mere stipulation. But Glock does go on to provide further indications of what it is that he believes separates behaviour from 'the information processing and activities of plants and micro-organisms' (ibid., 649):

> [. . .] our notions of behaviour and perception constitute a *Gestalt*. They *intertwine* behavioural with *morphological*-cum-*anatomical* criteria. Animal behaviour is animate *activity along a line of orientation determined by the position and alignment* of sense organs. The well-known alignment of sunflowers towards the sun, by contrast, is the direct effect of differential growth of cells. Those cells process Shannon information—concerning the direction of the sun;

> yet they are not specifically geared to *informing the organism* of that direction.
>
> (ibid., 649)

There are a number of ideas in this rich paragraph that seem to need separating. One is the idea that our notions of behaviour and perception constitute a 'Gestalt'. What does this mean, exactly? One natural way of interpreting the claim might be as the denial of the idea that the attribution of behaviour and perception to an entity can be disentangled one from another. What is involved, Glock seems to be suggesting, is rather a holistic appraisal, in which an organism's purposive movements are partly registered as such on the basis of anatomical considerations; and anatomy is decoded, in turn, partly by observation of behaviour (think of how we know which part of a small flying insect is the head merely by watching its motions, without being able directly to observe distinct anatomical features). A second idea in this paragraph is the suggestion that such plant motions as the movement of sunflowers to face the sun are the 'direct' effect of differential growth of cells—the idea perhaps being that where we have true behaviour, by contrast, effects are *mediated* in a distinctive way. And finally, we have the (perhaps closely related) idea that a plant, such as a sunflower, though it responds in a certain way to sunlight by turning towards it, is not itself *informed* by its cells (nor, presumably, by anything else), of where the sun is. What should we make of this intriguing collection of suggestions? Will it suffice both to found an acceptable distinction between behavers, on the one hand, and mere information-processors on the other—and is it clear that plants fall, as suggested, on the information-processing side of the divide?

To take *Gestalt* first, I think it may well be true that whether we are inclined to regard an organism as capable of perception is probably inextricably entangled with the question as to whether we are inclined to regard it as capable of behaviour. The question is, though, why we should not regard plants as both perceivers *and* behavers, just as we do animals—and whether the particular morphological-cum-anatomical criteria and locomotion-involving cues we seem, in fact, to rely upon to attribute both capacities represent anything more than evolutionarily helpful perceptual and cognitive biases that help us swiftly notice such things as potential prey and predators, as well as conspecifics, and assess their likely informational states (e.g. 'has it/have they seen me'?), within our own ecological niche. How are we to be confident that our intuitive exclusions and inclusions from the category of perceiver-behavers are not reflective of a narrow-minded preference for beings which have sense organs we can recognise as being at least a bit like ours—and/or which operate at timescales and in environments that have enabled us to observe and process their movements in ways that facilitate our interpretation of them as purposive, intentional, or even rational? How are

we to go about attempting to weed out mere biases that result from the contingencies of our particular perceptual mechanisms?[6] Someone keen to defend the idea that plants behave might protest that the basis of the Aristotelian distinction between the type of *psyche* belonging to animals and that proper to plants simply embeds a deeply problematic and partially hard-wired zoocentrism, and that were we able to observe the movements of plants at the kind of timescales which make the directed nature of their motions more evident, as we now can thanks to time-lapse photography, we might take a different view.[7] Anthony Trewavas, a prominent defender of plant cognition and intelligence, points out that watching time-lapse videos of plant motions tends to generate a strong conviction in observers that the motions involved both in growth and in other kinds of local movement of leaves, limbs, etc. are purposive (Trewavas 2009). And so far as perception is concerned, the question must be why we are *not* to say that plants perceive? We have abundant evidence that they respond to a wide range of environmental information in order to adjust their growth in various ways to the conditions in which they find themselves. What more would be required? And while it must be conceded that *locomotion* is not found in plants, it might be questioned why any special importance should be attached to the capacity to move from place to place, so far as behaviour is concerned. Does this not simply constitute an unreasoned prejudice against beings that are rooted to the spot? Perhaps plants may achieve, by way of resolving into a single determinate body shape the considerable possibilities afforded them by phenotypic plasticity, some of the very same sorts of things that animals achieve by locomotion? The question arises, therefore, whether the limitations of our own perceptual mechanisms and cognitive biases are distorting our view of what plants are, and what they do in such a way that we are, in fact, missing a good deal of what is there to be discerned in their slow-acting and distributed form of intelligence.

A very natural response to the question raised above about what more, beyond flexible responsiveness to environmental cues, is required for perception might attempt to invoke the thought that *true* perception (as opposed to mere information processing, perhaps) involves some form of conscious, sensory awareness of a kind widely supposed to be lacking in plants. Doubts might be raised about whether this really *is* an implication of the term 'perception'—I suspect myself that it is not—or not any longer, given the widespread use of the term in a varied range of scientific biological contexts. However, to fuss about that would lead us only into relatively unproductive arguments about semantics. If 'perception' is thought not to imply conscious sensory awareness, let us simply (for the sake of argument) replace 'perception' by 'conscious sensory awareness' in the characterisation of the other element of the 'Gestalt' that we take to be distinctive of the real behavers.[8] Might it be argued on Glock's behalf that since plants do not have such conscious sensory awareness,

they must lie on the opposite side of an extremely important boundary from many animals, one which deserves to be marked?

The difficulty, of course, is that while the boundary between the sentient and the non-sentient, the things which have conscious feelings, and those which do not, could scarcely be more important (tracking, as it does, one crucial determinant of whether a particular variety of *moral* concern is due to a given kind of being), there is no direct way of ascertaining that plants do *not* have conscious sensory awareness. As H.S. Jennings, the American zoologist commented, 'It is clear that objective evidence cannot give a demonstration, either of the existence or non-existence of consciousness, for consciousness is precisely that which cannot be perceived objectively' (2006, 336). It is true that we have been very apt to *suppose* that consciousness requires such things as a brain and nervous system—but there have been many recent attempts to question the assumption that no nervous system exists in plants—and even to suggest that certain structures within plants—the meristems—may be conceived of as possessing some of the same sorts of functions as brains serve in the integration of information. Calvo et al. (2020) point out that detailed experimental evidence was provided as long ago as 1926 for a plant nervous system with action potentials by J.C. Bose (Bose 1926), and recent work on plants continues to reveal hitherto unsuspected levels of complexity and sophistication in the ways they respond to their surroundings. On the basis of a theory of consciousness first proposed by Tononi (2004; 2008) and later developed in conjunction with, and also separately by Koch (Koch and Tononi 2008; 2011; Tononi and Koch 2015; Koch 2009; 2013; 2014; 2018), Calvo et al. (2020) argue that we should endorse the view that plants are conscious. That theory is known as the 'Integrated Information Theory' (IIT), and it proposes that consciousness requires a single integrated entity with a repertoire of highly differentiated states. Integration is key: consciousness, according to the IIT, will turn out to be a matter of degree, in a sense which can be very precisely quantified—it will depend on the extent to which an integrated system generates information over and above the information that is separately generated by its parts (with information classically defined as the reduction of uncertainty). A human conscious experience of seeing a scene in front of one, for instance, rules out billions of alternative scenarios at once, integrating, as it does, strictly visual information with a vast array of other contemporary sensory inputs, as well as memories, learned associations, expectations, and beliefs, which help categorise and order sensory information in ways that leave their stamp on the character of the experience itself.[9] The informational richness of such a visual experience is immense. More local information which occurs earlier in visual processing such as e.g. that possessed by groups of cells in the retina or the lateral geniculate nucleus, reduces uncertainty to a far smaller degree—and crucially, the separate informational contributions

of the various different local structures sum to significantly less than the informational upshot of the integrated whole. According to Tononi and Koch, it is this integration of information which is, as it were, the objective correlate of conscious experience.[10]

Tononi and Koch are neuroscientists, both seeking an account of full-blooded, qualitative, sensory consciousness in humans, and though they briefly consider possible applications of IIT outside the human context, they are mainly concerned with what it is that enables the *cerebral cortex* to achieve the feat of producing conscious experience. But Tononi is clear about the implications of IIT for the question how we should decide whether or not other beings are conscious:

> The IIT has a straightforward position on this issue: to the extent that a mechanism is capable of generating integrated information, no matter whether it is organic or not, whether it is built of neurons or silicon chips, and independent of its ability to report, it will have consciousness.
>
> (Tononi 2008, 237)

It is on the basis of this theoretical stance, together with the results of empirical investigations into the organisation of information processing in plants, that Calvo et al. argue that there are reasons to suppose plants have consciousness, albeit to a lesser degree than that likely to be found in many animals.

I cannot here undertake the considerable task of assessing either the IIT itself or the evidence that plants do indeed integrate information (though I shall describe a little of that evidence shortly, in dealing with Glock's idea that plants respond 'directly' to stimuli) and that on that basis, they may be considered conscious—though I find the idea highly suggestive and intriguing. My present point is merely the dialectical one that one cannot expect to get anywhere much in *justifying* the position that plants do not behave on the grounds that behaviour and conscious awareness are standardly co-ascribed according to a 'Gestalt' which intertwines morphological with behavioural criteria. For the defender of plant consciousness will simply want to know how we can be sure that those criteria are the right ones to supply the foundation for a scientifically adequate conception of behaviour, rather than mere spin-offs from an anthropocentrism that pervades our thought and experience, at both the personal and sub-personal levels, an anthropocentrism from which we ought to be attempting to struggle free, rather than embracing as the basis of our philosophy of behaviour.

What about Glock's second and third suggestions?—(ii) that plant motion is 'directly' rather than mediatedly controlled; and that (iii) no organism is 'informed' in the course of the information processing that plants undertake? Unfortunately, the same accusations of zoocentrism

seem potentially applicable here too. Take mediation first—is it true that where movement occurs in plants, the relevant effects are 'direct' in some way, which contrasts with some kind of mediatedness that we know to be present in animals? Certainly, there is a good deal more complexity in plant responsiveness to environmental factors than Glock's dismissal of the Shannon information processing that he takes to be involved in the orienting of sunflowers toward the sun would suggest. Trewavas, for instance, stresses that although some forms of purposeful behaviour in plants seem overwhelmingly to be controlled by one signal (the example given is the great sensitivity to unidirectional blue light in etiolated seedlings, which overrides opposing gravity signals), most plant responses involve the integration of numerous different pieces of information (Trewavas 2009). This knowledge goes back to Darwin (1880), who showed experimentally how the roots of seedlings integrated signals involving touch, light, moisture, and gravity, and that roots were able to 'decide' what weightings to permit to each. Moreover, it has been proven in more recent years that plants are able to make use of a much wider range of information than Darwin ever suspected. Plants are now known to respond to information concerning their own state of health; the presence or absence of conspecifics in the neighbouring area; volatile chemicals which may signal the presence, variously, of herbivores, parasites, damage to neighbouring plants, or the closeness of potential host plants (for parasitic species); the soil volume in which they are growing; and the spatial dimensions—both actual and predicted—of nearby competitors. Trewavas describes plants as 'compiling' information and with a range of inputs as complex as those listed above, there would seem to be nothing particularly 'direct' about the process by which they settle upon a suitable response. Moreover, these points are obviously connected also with the possibility of mounting an argument against Glock's claim that no organism is *informed* of anything during the process whereby a sunflower turns its face towards the sun. Whether or not this is the right thing to say about this particular piece of sunflower behaviour, it does seem as though there are certainly other plant behaviours which seem to involve a truly holistic top-down variety of responsiveness; for instance, if a whole shoot or root is cut from a plant, further plasticity changes are inhibited until the lost parts are regenerated—it is rather as though the whole plant 'knows' what has happened and responds accordingly. Perhaps Glock would still insist that talk of the plant's being 'informed' could never be more than a *façon de parler*; but it is an important question why exactly we should *not* make the literal claim—and why our extant inclination not to do so, even if we have one, should be given any legitimising weight.

In short, then, I am doubtful that there is much mileage in Glock's attempt to found a distinction between the perceiver-behavers, on the one hand, and the mere information-processors, on the other, in the mix of

considerations he sets forth. So far as behaviour is concerned, a continuum of complexity appears to be the order of the day—and in particular, it seems difficult to argue for a well-founded distinction which clearly excludes plants from the behaviour category, on the kinds of grounds that Glock suggests. The next question I should like to address, though, is whether we must say much the same sort of thing (*mutatis mutandis*) about the distinction between behaviour and *agency*—that is, that on inspection, it cannot be turned into anything respectable, and that we must settle for a *continuum* that admits of many differences of degree, but where there are no notable demarcations of the kind that are worth attempting to define. Here, I shall try to suggest, a negative answer may be easier—and more important—to defend.

From behaviour to agency

Glock's view on the presence or absence of agency amongst the lower animals, in plants, and in other simple forms of life is less clear than are his views about the 'higher' animals. In his (2019), he characterises the view that '*bona fide* agency equals rational or at any rate intentional agency' as a 'dogma' (645), and thereby certainly suggests that he is sympathetic, in principle, to the idea that *bona fide* agency extends somewhat beyond the realm of the animals he considers to be both rational and intentional agents. But how far, exactly? What is to be the test, for Glock, of *action*, as opposed to mere behaviour? At one point, Glock remarks that although we perhaps speak more readily of animal *behaviour* than of acts animals perform, the situation looks different once we turn our attention from the generic notion of action and consider various specific action verbs. Glock observes that:

> Animals do many of the things we do. They seek out food by foraging, hunting and killing prey; they eat; they court and mate; they protect their offspring; they bond; they defend their territory; etc. Some of them also play, prepare and hide food, make and use tools, socialize and act in concert, form alliances, fight and make up, etc.
> (Glock 2019, 650)

But, of course, not all animals do all these things—indeed, eating seems to be the only activity on his list that is, in some sense or other, universal amongst animals.[11] Perhaps, then, it is Glock's view that it is the capacity to engage in a certain special set of activities, of which the list he offers provides examples, which characterises the agents?

Of course, there are serious questions about how any list of this kind is to be extrapolated from Glock's indicative one, and also about whether any problematic anthropocentrism or other kinds of bias might be built into the choice of activities mentioned. There would also be worries

about how we are to distinguish the relevant usages of the set of favoured action verbs from what, I take it, would have to be regarded by Glock as misleading extensions to mere behavers, or even to information-processors, of the terminology proper to agency. Plants, for example, are described as 'foraging' for resources by many botanists; some have been characterised as protectors of their offspring (Robinson 2014). Glock, presumably, would need to insist that these usages are picturesque and non-literal applications of the relevant action verbs. But the list might at any rate give us a starting point from which to begin to think about agency as a capacity that might not extend right down to the midges, mussels, and millipedes, let alone the mosses, marigolds, and magnolias. For further clues to how we might proceed from this very rough and ready starting point, though, we must look more systematically at Glock's overall picture.

Glock sets out by noting two varieties of 'decoupling', which, he claims, might potentially be relevant to the distinction between agency, on the one hand, and mere behaviour, on the other. Some behaviour, he notes, is guided by perception in a way that is immediate and rigid. For example, a moth in a room in which there is a concentrated source of ambient light will fly immediately to the light[12] and will continue fruitlessly to fly into the bulb, even after experience might (one would have thought) have revealed this to be an unrewarding activity. This behaviour therefore reveals itself to be an unwanted side-effect of a light-responsive navigation system that does not have the flexibility to adapt to circumstances in which it is no longer fulfilling the useful purpose for which it evolved. But other kinds of animal behaviour show considerable flexibility as regards the question of when, how, and whether perceptual input of a relevant kind is to be acted upon—behaviour is 'decoupled', to use Glock's term, from any immediate stimulus. One very obvious kind of flexibility which seems relevant to agency is the capacity to choose *when* to act—a capacity which seems related to the ability to prioritise amongst competing goals, to plan for their *combined* attainment, and to attempt to achieve certain ends only when circumstances seem maximally propitious. Another is the capacity to take different means to a single end, perhaps even experimenting until the best method is found; or to take different routes to a place. All these competencies disrupt any straightforward and rigid connection between sensory input and behavioural output—and all are varieties of disruption which seems relevant to the question whether what we are observing is action, or is rather to be consigned to the category of mere behaviour, insufficiently within the control of the whole animal, as opposed to the programmes with which it comes ready-equipped, to count as agency.

The second variety of 'decoupling' Glock discusses is different. The explanation of some creatures' activities, he notes, requires a differentiation between cognitive states, such as belief and conative states, such as

desire and intention. As Glock notes, there are animals that pursue goals guided by perception which do not display this difference. He gives as an example the tendency of leopard frogs to snap at anything that looks like an insect, quite irrespective of its own state of satiation. Millikan argues that for such behaviours as that of the frog, we might wish to make use of the notion of a pushmi-pullyu representation which has neither (or both) of the classic 'directions of fit', being both imperative and indicative at once (Millikan 2005). But other creatures, as Glock sees it, are distinctive in having *wants*, states that are both distinguishable from mere pushmi-pullyu representations in having their own distinctive direction of fit, and which are, moreover, something more than mere concomitants of goal-directed behaviour *as such*. As Dennett has observed, we can apply an 'intentional stance' to more or less anything involved in goal-directed behaviour, and so we can attribute wants in a *sense* even to such things as thermostats and chess computers, as well as to many (all?) simple biological creatures. In this undemanding sense, the thermostat can be said to 'want', for instance, to keep the room at 19 degrees Celsius and that the chess programme 'wants' to get its queen out early (Dennett 1971; 1987). But Glock is after a much richer notion of *want* than this; and ultimately, it is in terms of this richer notion of a want that he chooses to characterise (what he calls) the power of agency:

> [. . .] some animals are capable of doing what they want to do, *pursuing goals that are not simply dictated by and may downright conflict with their biological needs* [. . .]. I call such flexible behaviour 'action', a terminological choice that is not dictated by yet perfectly compatible with established parlance.
>
> (2019, 651, my italics)

Glock thus ultimately chooses the second rather than the first variety of 'decoupling' that he considers as his criterion of the power of agency—his considered view seems to be that it is where we find wants of a kind that are not simply immediate, derivative attributions from the mere pursuit of biological requirements that we find this distinctive capacity to *act*.

I am happy to agree, with Glock, that the capacity to have wants that outrun and perhaps sometimes even conflict with biological needs is a philosophically significant capacity that deserves to be marked with some sort of label.[13] And Glock is entitled, of course, to make the terminological choices he wishes—he can call this significant capacity 'agency' if he likes. But I would question whether everything he says on the topic of agency is now consistent, given this terminological choice. Recall that Glock appears to wish to question the running together of intentional and/or rational agency with basic agency (recall that he calls this running together a 'dogma'). But has he not himself now engaged in exactly the same conflation in endorsing the possession of wants that

outrun mere biological needs as his chosen mark of agency? For the capacity to have wants that are not simply dictated by biological needs seems to me to be *already* a characteristic of agents that would have to be regarded as intentional and capable of acting for reasons, by Glock's own lights. Is it possible to imagine agents which could be regarded as *doing what they want* in Glock's strong sense (a sense in which the *mere* pursuit of biological goals does not count, in and of itself, as a generator of wants) yet not at the same time acting intentionally (and indeed 'for reasons', the reasons supplied by those wants?)? I cannot myself see how this combination of attributions could ever possibly arise. And if I am right that it could not, we must ask Glock where, then, is the room in his picture for the 'basic' agents, the ones we should be able to recognise if we are not to fall foul of the inflationist dogmas against which he warns us? Moreover, Glock's original list of action verbs certainly characterises a set of activities that are engaged in by creatures, which I doubt it would be appropriate to describe as having wants that outrun their biological requirements—spiders kill prey, mate, and fight, for instance; and wood roaches care diligently for their eggs for up to three years before they hatch and hence, it seems 'protect their offspring'. How is Glock's suggestion that we might in some way use such a list as a baseline from which to begin to think about agency to be reconciled with his much more stringent terminological direction to regard as agents only those creatures whose wants outrun the mere dictates of biology?

The suggestion I should like to make in order to conclude this chapter is that actually it is the first and not the second of the two kinds of 'decoupling' that Glock discusses that offers the more promising resources for a delineation of basic agency. Basic agency begins where creatures are freed, at least in some respects and to some extent, from the operation of programmes that cannot be switched off, or adapted, or deployed later rather than now, where that seems more advantageous; or deployed in this way rather than that, where *that* seems more advantageous, etc.—and where different behaviour can result on different occasions, even when circumstances seem similar. Where we have this combination of factors, we have some reason for saying that what happens is *up to the organism*, as opposed to produced causally by certain of its subsystems, parts, or genetically imprinted programmes. In his (1978), Harry Frankfurt famously pointed out that the movements of a spider when it moves its legs as it makes its way along the ground are correctly *attributable to the spider* in much the same sort of way, it would seem, as our movements are attributable to us when we move at will. We need to capture what it is in virtue of which it is right to say that it really is *the spider* (and not processes within it, parts of it, or programmes it comes equipped with) that is making the movements.

Agency thus conceived of is a capacity which plausibly is possessed only by some, but not by all behavers—for example, it does not seem to

me that paramecia are agents, judged by this standard, as I have argued elsewhere (Steward 2012). Although paramecia respond to a range of cues, they appear to do so in such a way as to generate stimulus-response relationships which are modellable by way of a set of fairly simple equations (Glaser 1924; Chase and Glaser 1930), so that the hypothesis that anything is left to the organism's *discretion* at the time of action seems extravagant. I have become much less confident about plants, because recent discoveries have made it seem plausible that we are only just beginning to make a start on understanding quite how sophisticated they may be. It seems possible to me, at any rate, that *some* plants might be agents, even while some simple multicellular animals are not—and we might need to let go of the idea that any plant ranks 'below' any animal in some fantasy chart that depicts the order of nature. But even if it turns out that no plant behaviour ultimately qualifies as agency, we need to be able properly to recognise that the top-down variety of control in which agency consists is not restricted only to the fairly narrow class of (probably mainly social) agents to whom wants which outrun biological needs may be ascribed. We need, in my view, to be able to see included in the category of agents the birds, many small mammals, and large herbivores; a large range of reptiles and arachnids; some fish; and doubtless many other kinds of animal which might fail the test Glock proposes. Glock's want-based criterion does not seem to me readily to allow for such broad membership of the agency club. It is in the development of ideas relating to the other kind of decoupling, in my view, the decoupling of response from stimulus, and powers such as those of waiting, planning, and choosing which have emerged in the space made available by this decoupling, where we are more likely to find the seeds of basic agency.

Having disagreed with the details of Glock's views, though, I want to end on a note of agreement. I endorse many important aspects of Glock's position, including the following three important claims: (i) There are many agents that do not possess language; (ii) we should resist the inflationism that collapses all agency into the intentional and/or rational varieties; but (iii) for all that, we need to retain the ability to insist that agency is something more complex and sophisticated than mere behaviour. In my view, though, this additional complexity is dependent on the ontological emergence, as life develops, of integrated organisms with distinctive powers of discretionary input into their environments; and does not require anything so high level as wants that outrun and even conflict with biological requirements. If we are to recognise Frankfurt's spider as an agent, we shall need to set the bar lower.

Notes

1 For a range of representative work, see Glock (2000; 2006; 2009; 2012; 2013; 2019).

2 Glock offers as examples of 'higher animals' the great apes, certain marine mammals, elephants, some avian species such as parrots and crows, and some cephalopods.
3 McDowell's claim that 'movements of limbs without concepts are mere happenings, not expressions of agency' (1996, 89) is a good example of an inflationary conception of agency. Other notable manifestations of the view are found in Davidson (1975; 1982), Hacker (2007), Stoecker (2009), Brandom (2010), and Marcus (2012).
4 Amongst these reasons might be, for example, the thought that (i) beliefs require fine-grained concept-possession of a sort that only language makes available (Stich 1979); (ii) the idea that intentions require that one 'know what one is up to' (Anscombe 1957); or (iii) the thought that intentional agents require a sense of the future and an ability to plan (Bratman 1987).
5 For a range of relevant ideas, see Brembs (2010); Hall (2011); Mancuso and Viola (2015); Linson and Calvo (2020).
6 Neuroimaging studies have revealed that humans respond more strongly to the observation of human, as opposed to non-human forms of movement. The observation of humanoid robots is sufficient to trigger the same strong response (Oberman et al. 2007). Our perceptual systems clearly already set up immediate biases of attention, interest, and interpretive ease; the question is whether in doing so, they serve effectively to conceal from us the workings of life-forms whose behavioural complexities we have had much less evolutionary need to probe.
7 Cf Pfeffer (1906, 2): 'The fact that in large plants, the power of growth and movement are not strikingly evident has caused plants to be popularly regarded as still life[. . .] . If mankind from youth upwards were accustomed to [. . .] perceive the activities of weeks or months in a minute as is possible by the aid of a kinematograph, this erroneous idea would be entirely dispelled'.
8 It is notable here that 'perception' may not in any case be the most accurate translation of the Greek notion of 'aisthêsis', which is used by Aristotle in his characterisation of the powers proper to animals—indeed, the translation by J.A. Smith which is incorporated into Jonathan Barnes's (1984) *Complete Works of Aristotle* goes for 'sensation', which to my ear already implies the suggestion of conscious feel.
9 Illusions such as the rubber hand illusion, the McGurk illusion, and others are very powerful proofs of the fact that expectations derived from one sensory modality can entirely change the sensations experienced in another.
10 It is important to be clear that the IIT is not suggesting that in view of the impossibility of ascertaining for sure which creatures possess conscious experience, we should *reconceive* consciousness for purposes of scientific tractability, as a property that can be defined informationally (in the same sort of way as those philosophers who developed the notion of 'access consciousness' (Block 1995) were sometimes accused of having reconceived it). Tononi and Koch are clear that they endorse the idea that consciousness involves the possession of genuine 'what-it-is-like-to-be-ness' (Nagel 1974). Their position is that 'what-is-it-like-to-be-ness' comes in degrees and arises out of the different degrees to which organisms engage in informational integration.
11 Though there are varieties of 'eating' that I—and I suspect also Glock—would not want to characterise as active. Glock himself discusses the leopard frog, which snaps at anything that comes within reach that looks sufficiently like a fly, quite independently of its state of satiation. That looks like programmed behaviour—the very thing from which Glock is trying to separate real agency.
12 It is this sort of behaviour that Dennett famously characterised as 'sphexish', a term coined from the Latin name for the digger wasp, *Sphex ichneumoneus*, some of whose rigid behaviour Dennett discusses in his (1984).

13 Sterelny (2003) offers a very illuminating discussion (Chapter 5) of the emergence of what he calls 'preferences', which seem similar to Glock's 'wants', and of the evolutionary conditions that may have made this emergence possible.

References

Anscombe, G.E.M. (1957). *Intention*. Oxford: Blackwell.

Barnes, J.A. (1984). *The Complete Works of Aristotle*. Princeton, NJ: Princeton University Press.

Barron, A.B. and Klein, C. (2016). What Insects Can Tell Us about the Origins of Consciousness. *Proceedings of the National Academy of Science of the United States of America* 113(18): 4900–4908.

Bassler, B. (2018). How Bacteria 'Talk'. https://evo2.org/intelligent-bacteria/. Last accessed: 29 March 2022.

Block, N. (1995). On a Confusion about a Concept of Consciousness. *Brain and Behavioural Sciences* 18: 227–287.

Bose, J.C. (1926). *The Nervous Mechanism of Plants*. London: Longmans.

Brandom, R. (2010). Conceptual Content and Discursive Practice. *Grazer Philosophische Studien* 81: 13–35.

Bratman, M.E. (1987). *Intention, Plans and Practical Reason*. Cambridge, MA: Harvard University Press.

Brembs, B. (2010). Towards a Scientific Conception of Free Will as a Biological Trait: Spontaneous Actions and Decision-Making in Invertebrates. *Proceedings of the Royal Society* B 278: 930–939.

Calvo, P., Baluška, F., and Trewavas, A. (2020). Integrated Information as a Possible Basis for Plant Consciousness. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/j.bbrc.2020.10.022.

Chase, A. and Glaser, O. (1930). Forward Movement of Paramecium as a Function of Hydrogen Ion Concentration. *Journal of General Physiology* 13: 627–636.

Chittka, L. (2017). Bee Cognition. *Current Biology* 27: R1049–R1053.

Chittka, L. and Wilson, C. (2019). Expanding Consciousness. *American Scientist* 107: 364–369.

Darwin, C. (1880). *The Power of Movement in Plants*. London: John Murray.

Davidson, D. (1975). Thought and Talk. In Guttenplan, S. (ed.), *Mind and Language*. Oxford: Clarendon Press; reprinted in Davidson's *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press, pp. 155–170.

Davidson, D. (1982). Rational Animals. *Dialectica* 36: 317–327, reprinted in Davidson, D. *Subjective, Intersubjective, Objective: Philosophical Essays*. Vol. 3. Oxford: Oxford University Press, pp. 95–106.

Dennett, D. (1971). Intentional Systems. *Journal of Philosophy* 68: 87–106.

Dennett, D. (1984). *Elbow Room*. Oxford: Oxford University Press.

Dennett, D. (1987). True Believers. In Dennett, D. (ed.), *The Intentional Stance*. Cambridge, MA: MIT Press, pp. 13–35.

Frankfurt, H. (1978). The Problem of Action. In *The Importance of What We Care About*. Cambridge: Cambridge University Press, pp. 69–79.

Gagliano, M. (2014). In a Green Frame of Mind: Perspectives on the Behavioural Ecology and Cognitive Nature of Plants. *AoB Plants* 7: 1–8. https://doi.org/10/1093/aobpla/plu075.

Ginsburg, S. and Jablonka, E. (2019). *The Evolution of the Sensitive Soul*. Cambridge, MA: MIT Press.

Glaser, O. (1924). Temperature and Forward Movement of Paramecium. *Journal of General Physiology* 7: 177–188.

Glock, H.-J. (2000). Animals, Thoughts and Concepts. *Synthese* 123(1): 35–64.

Glock, H.-J. (2006). Thought, Language and Animals. In Kober, M. (ed.), *Deepening Our Understanding of Wittgenstein*. Graz: Rodopi, pp. 139–160.

Glock, H.-J. (2009). Can Animals Act for Reasons? *Inquiry* 52(3): 232–254.

Glock, H.-J. (2010). Animal Agency. In O'Connor, T. and Sandis, C. (eds.), *A Companion to the Philosophy of Action*. Chichester: Wiley-Blackwell, pp. 384–392.

Glock, H.-J. (2012). The Anthropological Difference: What Can Philosophers Do to Identify the Differences between Human and Non-Human Animals? *Royal Institute of Philosophy Supplement* 70: 105–131.

Glock, H.-J. (2013). Animal Minds: A Non-Representationalist Approach. *American Philosophical Quarterly* 50(3): 213–232.

Glock, H.-J. (2019). Agency, Intelligence and Reason. *Philosophy* 94: 645–671.

Hacker, P.M.S. (2007). *Human Nature*. Oxford: Blackwell.

Hall, M. (2011). *Plants as Persons: A Philosophical Botany*. Albany: State University of New York Press.

Howard, S.R. et al. (2018). Numerical Ordering of Zero in Honeybees. *Science* 360: 1124–1126.

Jennings, H.S. (2006). *Behaviour of the Lower Organisms*. New York: Columbia University Press.

Koch, C. (2009). A Theory of Consciousness. *Scientific American Mind* 20: 16–19.

Koch, C. (2013). A Consciousness Meter. *Scientific American Mind* 24: 24–25.

Koch, C. (2014). Ubiquitous Minds. *Scientific American Mind* 25: 26–29.

Koch, C. (2018). What Is Consciousness? *Nature* 557: S9–S12.

Koch, C. and Tononi, G. (2008). Can Machines Be Conscious? *IEEE Spectrum* 45: 55–59.

Koch, C. and Tononi, G. (2011). A Test for Consciousness. How Will We Know When We've Built a Sentient Computer? By Making It Solve a Simple Puzzle. *Scientific American* 304: 44–47.

Linson, A. and Calvo, P. (2020). Zoocentrism in the Weeds? Cultivating Plant Models for Cognitive Yield. *Biology and Philosophy* 35: 1–27.

Mancuso, S. (2017). *The Revolutionary Genius of Plants: A New Understanding of Plant Intelligence and Behavior*. New York: Atria Books.

Mancuso, S. and Viola, A. (2015). *Brilliant Green: The Surprising History and Science of Plant Intelligence*. Washington, DC: Island Press.

Marcus, E. (2012). *Rational Causation*. Cambridge, MA: Harvard University Press.

Marshall, M. (2009). Why Microbes Are Smarter than You Thought. *New Scientist* 2009. https://www.newscientist.com/article/dn17390-why-microbes-are-smarter-than-you-thought/

McDowell, J. (1996). *Mind and World*. Cambridge, MA: Harvard University Press.

Millikan, R.G. (2005). *Language: A Biological Model*. Oxford: Oxford University Press.

Nagel, T. (1974). What Is It Like to Be a Bat? *Philosophical Review* 83: 435–50.
Oberman, L.M. et al. (2007). EEG Evidence for Mirror Neuron Activity during the Observation of Human and Robot Actions: Towards an Analysis of the Human Qualities of Interactive Robots. *Neurocomputing* 70: 2194–2203.
Pfeffer W. (1906) *The Physiology of Plants*, Vol. 3 (translated byA.J. Ewart). Oxford: Clarendon Press.
Robinson, M. (2014). Plants Can Actually Take Care of Their Offspring: Here's How. https://theconversation.com/plants-can-actually-take-care-of-their-offspring-heres-how-33048. Last accessed: 29 March 2022.
Sterelny, K. (2003). *Thought in a Hostile World*. Oxford: Blackwell.
Steward, H. (2012). *A Metaphysics for Freedom*. Oxford: Oxford University Press.
Stich, S. (1979). Do Animals Have Beliefs? *Australasian Journal of Philosophy* 57(1): 15–28.
Stoecker, R. (2009). Why Animals Can't Act. *Inquiry* 52: 255–271.
Tononi, G. (2004). An Information Integration Theory of Consciousness. *BMC Neuroscience* 5: 42.
Tononi, G. (2008). Consciousness as Integrated Information; A Provisional Manifesto. *Biological Bulletin* 215: 216–242.
Tononi, G. and Koch, C. (2015). Consciousness: Here, There and Everywhere. *Philosophical Transactions of the Royal Society* B370: 20140167.
Trewavas, A. (2009). What Is Plant Behaviour? *Plant, Cell and Environment* 32: 606–616.

13 Logic and the boundaries of animal mentality

Hanoch Ben-Yami

In the book, *The Hundred-Year-Old Man Who Climbed Out of the Window and Disappeared* by Jonas Jonasson, there occurs the following scene, involving the elephant Sonya:

> 'Sit, Sonya, sit!' said Allan as the final part of his daring plan.
>
> 'No, damn it, Sonya, *don't sit*,' shouted The Beauty, who suddenly realised what was about to happen [. . .].
>
> Sonya, who stood with her back to them all, had clearly and distinctly heard Allan's command. And the old man was nice to her, and she liked to do as he wanted. Besides, her benefactor and feeder [The Beauty] had *confirmed* the order. The function of the word 'Don't' to countermand an order was not something Sonya had ever grasped.
>
> So Sonya sat down.
>
> (Jonasson 2012, 122–123)

In a related passage, Wittgenstein reflects and suggests:

> One can imagine an animal angry, fearful, sad, joyful, startled. But hopeful? And why not?
>
> A dog believes its master is at the door. But can it also believe that its master will come the day after tomorrow?—And *what* can it not do here?—How do I do it?—What answer am I supposed to give to this?
>
> Can only those hope who can talk? Only those who have mastered the use of a language. That is to say, the phenomena of hope are modifications of this complicated form of life.
>
> (PI II.i.1)[1]

These and similar observations support the claim that there is an important gap between animals and us in basic capacities of understanding, feeling, and thinking. Such capacities will be the subject of this chapter. I shall relate some incapacities of animals to their lack of understanding of some basic concepts of logic.

DOI: 10.4324/9781003202929-17

Wittgenstein mentions above, *mastery of the use of a language*. Whether animals *do* have, in *some* sense, a language is, I think, a question too vague to have a clear answer. Animals do have calls whose function (at least in the sense that the eye has a function) is to alert others from their group to the presence of predators. They also express their emotions by vocalisations, something acknowledged from Aristotle on (*Politics* 1.2, 1253a9–1253a10; 7.13, 1332b5). Some animals have been trained to press on specific keys to get food and other things. Should we consider these, simple forms of language?—Call them as you wish. Some uses of language, I shall claim, involve the understanding of various logical concepts, which *we* master, whereas animals *do not*, and if so then such uses of language are beyond animals' capacities.

Negation. Children start saying *No* before they are two years old. No animal has yet signalled negation. To the best of my knowledge, there is no evidence of animal behaviour that shows a better understanding of negation than that which was exhibited by Sonya the elephant. (We shall presently consider experiments supporting this claim, and we will assess others that apparently refute it.) To master basic logic, one must understand negation. The inability of animals to use or respond to signs expressing negation shows they lack basic understanding of logic and whatever depends on it.

Whether animals can understand negation is an empirical question, not a philosophical one. In case it is discovered that some animals do exhibit such understanding, some claims below might not apply to them. Moreover, although we have a single symbol for negation in our formal logics, we can distinguish several kinds of incompatibility relations between statements. For instance, while 'Alice is standing' and 'Alice is *not* standing' are contradictory, 'Every student is standing' and '*No* student is standing' are contraries (they cannot both be true), and 'Some students are standing' and 'some students are *not* standing' can both be true but not false. It is possible that some kinds of relations we express using negation are understood before others, and that in this sense the understanding of negation is only gradually acquired. In that case, whether any animal has only a partial understanding of what we describe using negation should again be empirically decided. The relation I shall now consider, however, is that of a statement to its negation (p versus $\neg p$), and we shall see that animals have failed to show an understanding of this relation.

The understanding of negation is necessary for the understanding of the *disjunctive syllogism*: If you know that either A or B, and you later learn that A *is not* the case, you will infer that B. If animals do not understand negation, they should be incapable of inferring according to this rule. Do we have any evidence to the contrary?

In the third century BC, Chrysippus described a dog acting according to the disjunctive syllogism (Sextus Empiricus 1933, sect. I.69). This

anecdotal report, *if* it was meant as a report, cannot be trusted. Is there any *scientific* evidence for or against this capacity in animals?

In 2001, Watson and his collaborators published the results of the following study, conducted with dogs and with four- to six year old children (Watson et al. 2001). The dogs and children were first shown a desirable object in a container; next, a person holding the container passed behind three screens; and then, the container was shown to be empty. The dogs and children were then allowed to search for the object behind the screens. While children tended to increase their speed of checking behind the third screen after failing to find the object behind the first two, dogs tended to significantly decrease their speed of checking behind the third screen after failing in this way. We know that children of this age are capable of a disjunctive inference, and this explains their search pattern. The contrasting dogs' search pattern is explained if the dogs did not think logically but were motivated by mere association, and then each failure to find the object amounted to an extinction trial for the association.

If not dogs, what about apes? Redshaw and Suddendorf (2016) conducted the following study on children two- to four years old and adult chimpanzees and orangutans. 90 children and 8 apes were each given 12 times the opportunity to catch an item (ball for the children, grape for the apes) dropped into a vertical forked tube with two bottom openings. To secure catching the item, one should cover each opening with a hand, but not all participants have immediately or ever done that. Children's performance improved across age groups (2-, 2.5-, 3-, 3.5-, and 4-year-olds, 18 children in each), with none of the youngest group but most of the oldest group covering each opening with a hand the first time they prepared to catch the item. The 3- and 3.5-year-olds started with about 50% covering both openings, and after less than ten trials almost all of them covered both openings on at least some trials. About 50% of the 2- and 2.5-year-olds covered occasionally or consistently both openings after 11 trials. By contrast, of the 8 apes, only one covered both openings, and even that only once, during 12 trials. The apes' failures were not due to motoric limitations or inability to learn by trial and error: To establish that, the experiment was repeated with the apes many more times, and one of the eight, after covering both openings for the first time on the 15th trial, covered both of them for 33 more times out of its 69 later trials (Redshaw and Suddendorf 2016). I suspect the apes were more motivated to catch the grapes than were some of the children to catch the ball, yet none of the apes could realise that the grape could fall either from the right opening or *not* from it but from the left one and that therefore it should put a hand under each opening.[2]

It seems that observation and experiment support the conclusion of Mody and Carey (2016, 47): 'There is as yet no compelling evidence for successful logical reasoning using the disjunctive syllogism in nonhuman animals'. And, if animals cannot understand negation, then this

other inability follows necessarily. On the other hand, their failure to reason disjunctively provides empirical evidence for their inability to understand negation.[3]

This is not to say that animals cannot be *intelligent* in various ways, as Glock has emphasised (2019a). Chimpanzees have a short-term memory that surpasses ours, whereas elephants are said to have a good long-term memory. Some birds can learn complex sequences of action that lead to obtaining food. Some animals (chimpanzees, crows) use tools. Crows and other corvids adjust their food-hiding behaviour to the presence of other birds.—What *I* am trying to do, in a more Wittgensteinian fashion, is to consider which mental capacities animals do not and cannot have; and I try to derive these from their lack of understanding of negation and other logical operators.

An essential feature of our language is its *normativity*, namely, there is right and wrong in the use of words (Glock 2015, 2019b). We understand, for instance, that one has used a certain word *not* as it should be used, or that we do *not* yet know how to use it. Since animals do not understand negation, they *cannot* use signs in this way. We might train an animal to use a sign the way *we* intended it to be used, or the animal might not yet use the sign the way *we* intended; but the animal itself cannot understand that it does not know how to use the sign or that it has used it incorrectly.

Animals do not use signs the way we do. Some learn to use signs in a Pavlovian way, due to an association between previous uses and subsequent events. With others, the use of some signs is innate: East African vervet monkeys produce aerial alarm calls, distinct from those given in response to other predators; however, their close relatives, the West African green monkeys, who are not exposed to aerial predators and never use or hear such calls, also produce highly similar calls, distinct from their other calls, when presented with a drone (Wegdell et al., 2019). Accordingly, these signs involve no understanding, no right or wrong, and no communicative intention. We should be wary of ascribing to animals anything resembling our language.

The lack of understanding of negation affects the ability to understand other creatures. Animals can see that another animal is afraid, about to attack, that it sees something, or that it tries to get something, and they might modify their behaviour accordingly. But animals cannot understand that another creature is *wrong*: To ascribe a mistake to another creature is to ascribe to it a belief that is *not* true, and without negation such capacity is impossible. If animals cannot understand negation, then, necessarily, they cannot pass the false-belief task, if this involves understanding that another creature is wrong.[4]

And if you cannot understand that another creature is wrong, you cannot try to *deceive* others, namely try to make them believe something that is *not* the case:

> Why can't a dog simulate pain? Is it too honest? Could one teach (*lehren*) a dog to simulate pain? Perhaps it is possible to bring it (*ihm beibringen*) to howl on specific occasions as if it were in pain, even when it isn't. But the right surroundings for this behaviour to be real simulation would still be missing.
>
> (PI § 250)

> A child has much to learn before he can pretend. (A dog can't be a hypocrite, but neither can it be sincere.)
>
> (PI II.xi.363)

Unlike me, Wittgenstein does not try to relate this capacity to other, more basic capacities. I believe that here I am going beyond what is found in his writings.

Some animal behaviour might superficially appear as an attempt to deceive others. Tufted capuchin monkeys produce a sequence of a kind of hiccup calls in the presence of cats and snakes, which cause others in their group to become alert, look around for a threat, and occasionally escape right away. They seem to produce this call also under stress caused by other reasons. And some of them produce the call-sequence also in the presence of food, which often makes other monkeys retreat and increases their chances, especially if they are subordinate individuals, of getting food (Wheeler, 2009). Should we interpret this last case as *deceptive* behaviour? Should we even consider the call an *alarm* call? The production of the call *has* probably evolved because it alerted other monkeys to the presence of predators, and this makes it legitimate for *us* to classify it as an alarm call, but this does not amount to the monkey's producing the call because *it* wants to alert others. And as Wheeler also writes, these calls 'may be common during feeding because, after having once produced [them] in this context, observed the reaction of neighbouring conspecifics and acquired food as a result, individuals associate call production with access to food' (2009, 3017). More is needed to ascribe an intention to deceive, or even alert, to the monkey.

Animals cannot understand that others might be wrong, but they can be wrong themselves. The dog is barking up the tree because it saw the cat running up and it thinks it is there, yet the cat has jumped to the other tree and is gone. But can animals also have *illusions*? If this dog's case counts as one, then they can, but I do not think we would normally consider it such, and we certainly do not count every mistake as an illusion. One illusion would be to take an image in a mirror for a real person; another, to mistake a fake fruit for a real one and bite into it. If a dog barks at a mirror image or bites into a fake bone, does it have an illusion, or does it merely respond to them in the same way it does to the real things? With us, we can *realise* that we have been wrong, and this introduces a distinction between the two options; but is there anything

in the animal's case that distinguishes between them? An animal might have evolved to behave in a certain way in certain circumstances, and we can make it behave the same way in different, contrived circumstances, and this behaviour might then be detrimental to it: Does the animal have an illusion then, or is it just another reason for the same behaviour?

The best response might be, these options *amount to the same thing* in the animal's case, although with us they are distinct. Also, if there were only animals, there would be no need for the concept of illusion, as one could make do with 'another reason for the same behaviour'. And explaining what illusion is only through animal behaviour will not give us *our* concept of illusion. We can say, *such* illusions (*if* we consider them illusions) are a two-dimensional projection of our three-dimensional concept, or a limiting case of ours (cf. PI §§ 49, 385, 419). What I have in mind resembles the following: An ellipse is a closed plane curve whose standard equation is, $x^2/a^2 + y^2/b^2 = 1$. Put a = b, and you get a circle, $x^2 + y^2 = a^2$. So a circle, in a sense, is an ellipse, and that is fine. But if we had only circles, we would not need the notion of an ellipse: The *need* for that notion, and not just for that of a circle, arises from the fact that we consider also cases in which $a \neq b$. And animal-illusion: human-illusion = circular-ellipse: ellipse.

Another notion of logic, apart from that of negation, that is essential for our reasoning and understanding is that of *possibility*. Some things are *not* the case, but they *can or could* be the case. The understanding of modal notions depends, in this respect, on the understanding of negation (although not reducible to it), and is, therefore, together with the latter, beyond the mental capacities of animals.

Morality is, therefore, also beyond the understanding of animals, as it involves the *normative* notions of right and wrong, of what one *can and should* do, and of what one *can but should not* do (closely related to the normative notions involved in *our* use of language).

Several studies have been conducted to show that animals *do have* moral notions and understand injustice. The one best known is probably that of Frans de Waal and his colleagues with capuchin monkeys.[5] One monkey gets furious when it continues to receive cucumbers after it sees the other monkey receiving grapes for the same task. However, the monkey gets upset not because it thinks it was treated unjustly, but because it expects grapes and receives cucumbers. The monkey *does not* get upset when it sees the other receiving a grape after it received a cucumber: The obvious injustice does not upset it, the way it would us. Rather, it gets upset only later, when it does not receive what it expects. It cries in frustration, not with moral indignation.

We should not immediately interpret behaviour that with *us* is a result of certain feelings or beliefs as resulting also in animals, in similar *narrow* circumstances, from the same feelings or beliefs. We should rather first examine the animals in other circumstances as well, to determine

the limits of their capacities. Same effect need not mean same cause—and in the capuchin monkeys' case even the effects are different.

Some *feelings and emotions* involve understanding of modalities. Hope is one such: We hope that something that is *not* the case but *can* be the case will happen. Hope is not a special kind of sensation, although it might be accompanied by typical sensations and physiological reactions (patterns of breathing, say). The understanding of modalities is beyond the capacities of animals, and with it is hope, as Wittgenstein observed. When you take the leash off the hook, your dog *expects* you to take it out for a walk and that is why it gets excited. But when you take a nap, the dog cannot *hope* that you will take it out once you get up.

Remorse is similar:

> Why can a dog feel fear but not remorse? Would it be right to say, "Because it cannot talk"?
>
> Only someone who can reflect on the past can repent. This does not mean, however, that *according to experience* only such a one is capable of the feeling of remorse.
>
> (RPP II §§ 308–309, my italics)

The last remark is meant as a conceptual observation, not as based on experience. One need not run experiments to determine whether unmarried men are bachelors. We feel remorse if we think we *could have* done something different and better than what we did. (Again, remorse is not a special kind of sensation.) A dog cannot think that, and it cannot therefore feel remorse. Moreover, it also cannot *reflect on the past*—let us consider animal *memory*.

Animals remember many things: Animals in the wild remember where food and water are found, the dog remembers the trick it was taught yesterday, and more. But can the dog remember *being taught the trick yesterday*, namely, can it have *episodic memory*?

If *A* happened before *B*, then when *A* happened *B* has *not* yet happened. A thing that happened in the past *is not* or *might not be* the case now. The future event *has not* yet happened. Understanding the temporal order of events and their being past or future requires the understanding of these relations.

> We say a dog is afraid its master will beat it; but not: it is afraid its master will beat it tomorrow. Why not?
>
> (PI § 650)

> "So someone who has not learned a language is unable to have certain memories?" Of course, – he cannot have linguistic memories, linguistic wishes or fears, and so on.
>
> (PI § 649)

Episodic memory is, therefore, beyond the intellectual capacities of animals.

Studies *have* been conducted to examine whether animals do have episodic memory. An influential one is that run by Clayton and Dickinson (1998), in which scrub jays were shown to change their cache recovery preferences according to the time elapsing from caching food and food type. In one experiment, they cached both non-perishable peanuts and their preferred, though as they had learned, perishable waxworms (caterpillar larvae of wax moths); they usually first recovered the waxworms only shortly after caching them, when still fresh, while when they could recover the cached food only after a few days, when they would expect the waxworms to have decayed, they usually first recovered the peanuts.

However, unlike Clayton and Dickinson (1998, 274), I doubt this shows that 'the cache recovery pattern of scrub jays fulfils the three, "what", "where" and "when" criteria for episodic recall'. It shows that scrub jays expect to find food of the kind cached where they cached it, not that they remember the episode of caching it there; and that their motivation to recover perishable food declines with time. The latter, which is the original contribution of this study, does not indicate remembering the caching event and estimating the time that has passed since its occurrence; it is rather like the decline of anger, fear, and other emotions over time. These emotional-motivational states decline with the distance in time from the event that caused them, a process that does not require remembering that event, and so does the motivation to recover decaying food.

The fact that animals do not have episodic memory is a consequence of their limited conceptual capacities. This is not the current approach in cognitive science: People there wonder, might animals have a *representation* of past events?—However, a representation is much like a picture (a logical picture), and carrying a picture with you, in your pocket or in your brain, does not mean you know what it depicts. Animals do not have the capacity to know what is depicted in such pictures, and therefore questions about their 'representations' are irrelevant.

The inability of animals to have episodic memories due to their cognitive limitations might be related to the phenomenon called *childhood amnesia*, the inability of older children and of adults to recall events occurring before the age of, usually, three to four. Children's cognitive capacities up to that age might be insufficiently developed to have a proper understanding of temporal concepts, which are necessary for episodic memory. Such an understanding is shown in the proper use and responses to language using different tenses, and the failure of children up to roughly the same age to pass the explicit 'false-belief' task might be a related phenomenon, also exhibiting a lack of understanding of tenses (Ben-Yami et al. 2019). It would, therefore, be interesting to check

whether there is a correlation between age in first memories and age of first passing that task. 'Childhood amnesia' would then mean not that young children *forget* but that they *cannot have* episodic memories. However, children in their early school years also *forget* early childhood episodes more than do older children, and children may have difficulties consolidating the memories that they do have, so childhood amnesia might comprise several distinct phenomena and be only partly due to the lack of cognitive capacities at an early age.

Lastly, can animals *dream*? A dream is 'a memory phenomenon of the awakened' (PI II, vii § 53): You wake up and you *seem to remember* having taken part in various events. (The physiological reasons for this apparent memory, which might be varied and might overlap only a little with the physiological occurrences when really participating in such events, do not interest us here.) If so, then animals cannot dream, and for the same reason that they do not have episodic memory: They do not possess the conceptual resources necessary for such apparent memory. Their twitches and turns during sleep should be considered mere physiological occurrences.

Animals can suffer, enjoy, be angry, surprised, or afraid. Some are distressed when they lose their young. They can also be intelligent, in a variety of ways. In these and in other respects, they resemble homo sapiens. But there are also important differences between them and it. And here I have chosen to follow Wittgenstein, but also philosophers very different from him and from each other—Aristotle and Descartes—and explore how the possession of *logos* endows humans, noble in reason, with a pre-eminence above animals, which lack it.[6]

Notes

1 I occasionally introduce minor changes to the translations of Wittgenstein's work used here.

2 Lambert and Osvath (2018) thought they found a problem with the results of Redshaw and Suddendorf (2016), because the chimpanzees in their experiment not only showed the same behaviour as those of Redshaw and Suddendorf's in theirs when the future results were *uncertain* (disjunctive), but some also when they were *certain*: Two tubes were simultaneously used, the chimpanzees saw that to each tube a grape is dropped, yet only three out of six chimpanzees learned to use rather consistently both hands in such cases, and even that only after many trials.—But the chimpanzees that started with the certain results were only mildly motivated to cover both openings in these cases, since they anyway had a continuous supply of grapes as the interaction continued. Those that started with the uncertain, disjunctive trials, in which they failed, were more motivated to get both grapes, for they had experienced getting none, and indeed two of the three slowly learned to cover both openings (but this higher percentage might be unrepresentative). I think Lambert and Osvath only *corroborated* Redshaw and Suddendorf's results, showing that the chimpanzees act with little understanding of either situation.

3 The conclusions of the authors of the recent (Ferrigno et al. 2021) study, that baboons can reason disjunctively, I find unjustified by their results. At best, two of the nine baboons with which they started continued searching above chance level around the place where they had seen a single grape being deposited after they had not found it under one of two cylinders there (does looking around deploy disjunctive syllogism?), and this only 72% of the considered trials, with 8% of trials in which the monkey checked again under the same cylinder not taken into consideration, trials with no responses not taken into consideration, 18% of trials that were either noted as unclear by one of two coders or on which the coders disagreed not taken into consideration, with performance slightly worsening as testing progressed, and performance not being significantly above chance on trials in which the first cylinder chosen was baited.

4 Krupenye et al. (2016) claimed that great apes do anticipate that other individuals will act according to false beliefs, and from this de Waal (2016) was quick to conclude that "apes know what others believe." I think that even if their experiments' results are replicable (given the exact data, I doubt we have strong grounds to trust that), their claim, and even more so de Waal's excited conclusion, are not warranted by these results. At most, apes *associate* the thing with which an agent recently interacted with the agent, as I have explained in more detail in a letter published in *Science* following their paper (Ben-Yami 2016).

5 See the video at https://www.youtube.com/watch?v=meiU6TxysCg.

6 I presented earlier versions of this work at a Wittgenstein graduate seminar I taught at CEU; at the Festschrift Symposium for Glock, "Wittgenstein and Beyond: Language, Mind, and Normativity"; and at the colloquium of the Department of Cognitive Science, CEU; and I benefited from the responses at all these events. Peter Hacker also wrote helpful comments on an earlier version.

References

Ben-Yami, H. (2016). RE: Apes Still Don't Understand Much (Response). *Science* 354(6308). https://www.science.org/doi/10.1126/science.aaf8110. Last accessed: 30 March 2022.

Ben-Yami, H., Ben-Yami, M. and Ben-Yami, Y. (2019). What Does the So-Called False Belief Task Actually Check? https://doi.org/10.23668/psycharchives.2513.

Clayton, N.S. and Dickinson, A. (1998). Episodic-Like Memory during Cache Recovery by Scrub Jays. *Nature* 395(6699): 272–274. https://doi.org/10.1038/26216.

de Waal, F.B.M. (2016). Apes Know What Others Believe. *Science* 354(6308): 39–40. https://doi.org/10.1126/science.aai8851.

Ferrigno, S., Huang, Y. and Cantlon, J.F. (2021). Reasoning through the Disjunctive Syllogism in Monkeys. *Psychological Science* 32(2): 292–300. https://doi.org/10.1177/0956797620971653.

Glock, H.-J. (2015). Meaning and Rule Following. In Wright, J.D. (ed.), *International Encyclopedia of the Social & Behavioral Sciences*. 2nd edition. Oxford: Elsevier, pp. 841–849. https://doi.org/10.1016/B978-0-08-097086-8.63049-0.

Glock, H.-J. (2019a). Agency, Intelligence and Reasons in Animals. *Philosophy* 94: 645–671. https://doi.org/10.1017/S0031819119000275.

Glock, H.-J. (2019b). The Normativity of Meaning Revisited. In *The Normative Animal?* New York: Oxford University Press. https://doi.org/10.1093/oso/9780190846466.003.0015.

Jonasson, J. (2012). *The Hundred-Year-Old Man Who Climbed Out of the Window and Disappeared*. Translated by R. Bradbury. Hachette Books.

Krupenye, C., et al. (2016). Great Apes Anticipate that Other Individuals Will Act According to False Beliefs. *Science* 354(6308): 110–114. https://doi.org/10.1126/science.aaf8110.

Lambert, M.L. and Osvath, M. (2018). Comparing Chimpanzees' Preparatory Responses to Known and Unknown Future Outcomes. *Biology Letters* 14(9): 20180499. https://doi.org/10.1098/rsbl.2018.0499.

Mody, S. and Carey, S. (2016). The Emergence of Reasoning by the Disjunctive Syllogism in Early Childhood. *Cognition* 154: 40–48. https://doi.org/10.1016/j.cognition.2016.05.012.

Redshaw, J. and Suddendorf, T. (2016). Children's and Apes' Preparatory Responses to Two Mutually Exclusive Possibilities. *Current Biology* 26(13): 1758–1762. https://doi.org/10.1016/j.cub.2016.04.062.

Sextus Empiricus. (1933). *Outlines of Pyrrhonism*. Translated by R.G. Bury. Cambridge, MA: Harvard University Press (Loeb Classical Library, 273).

Watson, J.S., et al. (2001). Distinguishing Logic from Association in the Solution of an Invisible Displacement Task by Children (Homo sapiens) and Dogs (Canis familiaris): Using Negation of Disjunction. *Journal of Comparative Psychology* 115(3): 219–226. https://doi.org/10.1037/0735-7036.115.3.219.

Wegdell, F., Hammerschmidt, K. and Fischer, J. (2019). Conserved Alarm Calls but Rapid Auditory Learning in Monkey Responses to Novel Flying Objects. *Nature Ecology & Evolution* 3(7): 1039–1042. https://doi.org/10.1038/s41559-019-0903–5.

Wheeler, B.C. (2009). Monkeys Crying Wolf? Tufted Capuchin Monkeys Use Anti-Predator Calls to Usurp Resources from Conspecifics. *Proceedings of the Royal Society B: Biological Sciences* 276(1669): 3013–3018. https://doi.org/10.1098/rspb.2009.0544.

Wittgenstein, L. (1980). *Remarks on the Philosophy of Psychology* Vol. II [=RPP]. Edited by G.H. von Wright and H. Nyman. Translated by C.G. Luckhardt and M.A.E. Aue. Chicago, IL: The University of Chicago Press.

Wittgenstein, L. (2009). *Philosophical Investigations* [=PI]. 4th edition. Edited by P.M.S. Hacker and J. Schulte. Translated by G.E.M. Anscombe, P.M.S. Hacker, and J. Schulte. Chichester: Wiley-Blackwell.

14 Two notions of creativity

Julia Langkau

Conceptual tensions

According to a widely accepted definition originally provided by Margaret Boden (1990/2004), creativity is the ability to come up with ideas or artefacts that are new relative to some comparison class, that are surprising or unexpected, and that are in some way or other valuable, appropriate, or useful.

When thinking and theorising about creativity, some of our intuitions are seemingly in conflict with one another. Tensions occur in at least three different sets of intuitions: (a) intuitions concerning animal creativity, (b) intuitions concerning artificial intelligence (AI) generated art and inventions, and (c) intuitions concerning the value component in our notion of creativity. I will address each of these sets in turn.

a. Animal creativity. We often assume that creativity is a genuinely and uniquely human ability (e.g. Sawyer 2012), one that distinguishes us from other animals and has brought progress to everyday life and to human civilisations (Hennessey and Amabile 2010). In face of a fast-changing world and ever-growing challenges such as the global ecological crisis humanity is being confronted with, creativity is becoming more and more relevant, and educators agree that fostering creativity in children is crucial. Julian Manley (2020), e.g., has recently argued that creativity is one crucial means in our search for a radical solution to the current climate crisis. The Future of Jobs Report 2020 of the World Economic Forum lists creativity as one of the ten most important skills we need in order to be successful in the 'Fourth Industrial Revolution'. Prime examples of creative humans are scientists and artists such as Albert Einstein or Frida Kahlo. However, it also seems obvious that animals can be creative, and researchers show a growing interest in animals as intelligent actors (e.g. Glock 2019). So far, most creativity research on animals concerns primates and birds. Chimpanzees, for instance, have been found to explore new and inventive ways of using a tool (Tomasello 2000), and grey parrots can use their abilities to

DOI: 10.4324/9781003202929-18

solve problems that are significantly different from problems they were trained to solve; similar to what human children do when they learn a language, they can communicate needs or goals using vocalisations in creative ways (Pepperberg 2015).

b. AI inventions and art. Besides distinguishing humans from other animals, creativity is also often thought of as distinguishing human intelligence from AI. Computer programs may be superior to the human brain in many respects, but they cannot be creative, at least they cannot *really* be creative, since they are programmed by humans, and 'being programmed is the antithesis of being autonomous' (Boden 2014, 229).[1] Autonomy is often taken to be a necessary condition on creativity, as are consciousness, intentionality and imagination. While currently existing computers do not seem to have these features or abilities, future computers may one day be autonomous, conscious, etc. and thus be able to be creative. However, computer programs are already quite complex and have produced work indistinguishable from humanly produced art such as music and paintings (see Gangadharbatla 2021). Computer programs have also generated useful inventions, e.g., Steven Thaler's Creativity Machine has generated the cross-bristle design of the Oral-B CrossAction toothbrush (see Abbott 2016, 1085; similarly useful inventions have been made by computer programs imitating the process of biological evolution, e.g., the Invention Machine by John Konza). Drawing on past success and constant development of AI algorithms, Ryan Abbott (2016) claims that 'A creative singularity in which computers overtake human inventors as the primary source of new discoveries is foreseeable' (Abbott 2016, 1080). Some may find this development worrisome, because it threatens the idea that creativity is something uniquely human and that humans cannot be replaced by computer processes at least in some crucial and defining areas. However, the creativity of some computer-generated outputs cannot be denied.

c. Creativity and value. It has been more or less common sense that we need to distinguish 'original nonsense' (Kant 2000, 186) from creativity, which involves value, since not everything that is new and surprising is also valuable. But Alison Hills and Alexander Bird (2019) have recently argued for a more inclusive notion of creativity that does not involve value. Their argument goes roughly as follows. When we look at certain products that are the result of what looks like a creative process, some of them are valuable and others are not. Following the idea that creativity involves value, we judge that the ones that are valuable are creative, but those that are not valuable are not. However, the mental process involved in generating both kinds of products must have been more or less the same. Hence, creativity cannot involve value.

One way of dealing with any conceptual tension is to make our theory fit in one way or the other, and to brush off any remaining conceptual uneasiness for the sake of a coherent theory. Sometimes, however, it is worth exploring more radical possibilities of accommodating most or even all of our conflicting intuitions. The goal of this contribution is to do exactly that by pointing out the advantages of distinguishing two different notions of creativity.

I will argue that the source of tension in all three sets of intuitions is the fact that theories of creativity have focused on the ability of a subject to generate a certain output, while also trying to account for the intuition that creativity involves a certain mental process. In all three sets, there is a tension between our judgements concerning the creativity of a product, on the one hand, and our judgements concerning the creativity of a mental process, on the other hand. Keeping two notions of creativity apart and applying one to the process and one to the product of creativity will resolve the tension and account, among other things, for the rapid development of AI generated art and inventions. It will also explain why we are interested in animal creativity and in the creativity of children.

I will first present several distinctions that have been made in the literature and that provide some relief, but I will also point to some of their limitations (see the section 'Different types and levels of creativity'). I will then argue that for the purposes of understanding creativity, it is best not attributed primarily to subjects, but rather to mental processes and to products separately and in two different senses (see the section 'Attributing creativity'). The reason is that we care about two different sets of aspects concerning creativity: sometimes, we care about aspects concerning a creative process, and other times we care about aspects concerning a creative product. I will show how distinguishing two notions of creativity can explain our conflicting intuitions (see the section 'Process creativity and product creativity'), and I will give independent support for the idea of two different notions (see the section 'Process creativity and product creativity'). The chapter concludes with the claim that future research on creativity should keep process creativity and product creativity apart (see the section 'Monkey selfies and cross-cultural differences').

Different types and levels of creativity

Philosophers and psychologists have introduced different types and levels of creativity to account for creativity in everyday life, in the sciences as well as in the arts. Boden (2010, 29–40; also 1990/2004) introduces different *types* of creativity by identifying three different ways in which a new idea can be surprising. First, a new idea can simply be unfamiliar; second, it can be surprising but fit into a known conceptual space; third,

it can be new in the sense that it was previously thought of as impossible. These three ways in which a new idea can be surprising correspond to the following three types of creativity: 'combinational creativity' consists of an unfamiliar combination of familiar ideas, e.g., when certain ingredients are being combined in a cake for the first time; 'exploratory creativity' involves an exploration of conceptual space, e.g., the discovery of formerly unknown implications of a philosophical theory. The third type, 'transformational creativity', involves thinking or creating something within a new conceptual space, e.g., starting a new movement in art, such as impressionism or photorealism. In a similar way, psychologists James Kaufman and Ronald Beghetto (2009) have distinguished different types of creativity by introducing what is now widely known as the 'Four C' model of creativity, which accounts for very basic creativity ('mini-c'), everyday problem-solving creativity ('small-c'), ordinary inventive creativity ('Pro-C') and extraordinary creativity ('Big-C').

Boden further introduces two different *levels* of creativity on the basis of the question to whom the product is new, surprising and valuable: We are concerned with historical creativity when the creative product is new, surprising and valuable in human history, but psychological creativity only requires that the product be new, surprising and valuable to the creative subject. Other philosophers have distinguished more fine graded levels: Something can be new relative to a wider range of comparison classes, for instance relative to a particular place or time, or to a particular subject, etc. (Stokes 2016, 247). For instance, a child's drawing may be creative within the context of a certain age group or school class, but not in a wider context.

These distinctions, both of types and levels of creativity, can help us understand our intuitions in (a): they can explain how animals, children and people in everyday situations on the one hand, and individuals such as Albert Einstein or Frida Kahlo on the other hand can be called creative. In terms of types of creativity, animals can be creative in the sense of combinational or at most exploratory creativity. In terms of levels of creativity, animals and ordinary people in everyday situations can be psychologically creative: The ideas or products they generate are new and valuable to them, or relative to a certain comparison class. Individuals such as Albert Einstein or Frida Kahlo, on the other hand, were able to engage in historical creativity: They came up with ideas or products valuable and new in human history, and the type of creativity they were engaged in should be categorised as transformational creativity. And surely it is the transformational type of creativity that is needed to overcome major challenges humanity is facing, not the type of creativity that animals, children and people in everyday situations can engage in.

But Boden's distinctions are controversial and in need of further specification (see Novitz 1999; Stokes 2011; Gaut and Kieran 2018). I will discuss only two points. First, Boden's distinction between the two

levels of creativity, psychological and historical creativity, catches an important aspect of our use of the notion of creativity: We are willing to call a child's drawing creative (psychological creativity) while also acknowledging that there are very few truly creative people and products (historical creativity). However, the distinction is based on a factor that can be completely independent of the creative process and even of the product. Here is an example. Apparently, modern calculus was developed by Isaac Newton and Gottfried Wilhelm Leibniz more or less at the same time. According to Boden, arguably only one of their ideas can be called historically creative, because only one was new to humanity (the one published/invented first). But intuitively, Leibniz and Newton are both creative in exactly the same sense, even if they did not come up with or publish their ideas at exactly the same time. It is clear that the distinction and, in particular, psychological creativity serves to avoid the consequence that only the person who first comes up with an idea can be considered creative at all (see Boden 2014, 227), but this becomes implausible if the difference in time is marginal. Whoever came up first within a small time frame is completely irrelevant, especially in the case of major achievements such as Newton's and Leibniz'. I will show below that distinguishing process creativity and product creativity can account for cases like Newton and Leibniz' much better.

Second, Boden's three different types of creativity introduced above ('combinational', 'exploratory' and 'transformational' creativity) are based on the kind of surprise we experience when confronted with the creative product. Boden claims that '[p]henomenologically, there are three types of surprise that we may feel on first encountering a creative idea' (Boden 2014, 228). However, while surprise obviously comes in degrees, it is unlikely that it comes with a certain phenomenology concerning the structure of what surprises us. Moreover, how surprised a person is will depend on their individual psychological makeup and on relevant previous experience. We can easily find examples in which the level of surprise does not correspond to the type of creativity Boden's theory would predict. Take Ursus Wehrli's work titled 'Tidying up Art'.[2] In this work, Wehrli takes existing art and re-organises its shapes and colours, i.e., reassembles them in ways that are statistically informative or, as he puts it: 'tidy'. When we look at the way in which Wehrli's work is creative, it is a prime example of creativity as the re-organisation of existing elements and hence of combinational creativity. However, it can be experienced as something completely unexpected. In some sense, it seemed impossible to take art apart and re-organise its components, and we likely have never before looked at art in terms of statistical occurrence of shapes and colours. The surprise is very powerful and amusing. In fact, by mere re-organisation of components, we get to see things in a completely new way. Hence, while the distinction between different ways in which something is new is useful, nothing follows with respect

to the nature of our reaction to it or, arguably, with respect to the creative process. Conversely, nothing follows from the kind of surprise we experience when we encounter a creative work. Instead, the different types of creativity should be understood as structural features of the creative product only (see the 'Process creativity and product creativity').

We have seen that Boden's distinctions can help us resolve the tension in (a): animals can be creative in the sense of combinational or at most exploratory creativity, and animals and ordinary people in everyday situations can be creative in the sense that the ideas or products they generate are new and valuable to them. Individuals who we appreciate and admire for being truly creative are creative in the transformative sense, and their creative ideas or products are historically creative.

As I will show in the next section, Boden's account does not help us resolve the conceptual tensions in (b): it does not account for the fact that computer programs can already generate products that are new, surprising and valuable. The reason is that while Boden's definition requires judgements of the product of creativity primarily, she claims that the kind of product, an idea or an artefact, implies production in a certain way, which concerns the mental process of creativity. This, I will argue, is a mistake, and it blends together two different notions of creativity. Even though the relevant phenomena often occur together, we need to distinguish two notions, and doing so not only avoids certain problems with Boden's distinctions, it can also resolve the tensions in all three cases.

Attributing creativity

We can call a subject, a product, or a process 'creative' (see Stokes 2016), and philosophers have disagreed about which of them is or are (more) basic to explaining the phenomenon of creativity. Boden's definition is concerned with the ability of a subject to generate a certain output and hence implicitly with how we judge this output: The idea or artefact has to be novel, surprising, and valuable. Clearly, computer programs can already generate such products, especially computer programs that use genetic algorithms, which can make random changes in the programme's own rules (Boden 2014, 229ff). In any case, if we take creativity to be the ability to generate a certain output, whether such computer programs can be creative seems to be an empirical question: We simply take a look at the products and decide whether they are new, surprising and valuable. But in her paper 'Creativity and Artificial Intelligence: A Contradiction in Terms?', Boden names several features of computer processes that may be creativity-denying, and several features computer processes lack that may be necessary for creativity. Requirements concerning the process of creating enter Boden's definition through the notion of an idea, e.g.: 'My own definition of creativity [. . .] tacitly assumed that the

novel idea was freely generated by the person concerned' (Boden 2014, 233). (The same applies to a novel artefact.) Adding requirements on the process of creating turns the question of whether computers can be creative into a philosophical question (Boden 2014, 224): Can computers *really* create? Boden's answer is that currently, they cannot—but given that AI is becoming more and more complex, they might one day have the relevant features.

In this section, I will look at how psychologists and philosophers attribute creativity and what this tells us about current theories of creativity, and what the focus of future research in philosophy should be.

Psychologists are interested in measuring creativity; they study the phenomenon as part of a person's psychological makeup, with respect to the environment, the psychological context, and the output all together. We can distinguish two different approaches (see Sternberg 2006). First, some psychologists focus on the cognitive process of creating in order to operationalise creativity. The *Torrance Test of Creative Thinking*, based on Guilford's (1956) theory of creativity, is one way of measuring creativity. According to Guilford's theory, creativity is divergent thinking and can be operationalised using four aspects that are individually testable: flexibility, defined as the number of categories used when solving a problem; fluency, defined as the ability to produce a large number of relevant ideas; originality, defined as the ability to produce ideas that are statistically rare; and elaboration, defined as the ability to implement an idea in detail and high quality. Second, some psychologists focus on aspects in the psychological makeup that have to be in place in order for creativity to emerge. For instance, Amabile's (1988) theory of creativity claims that creativity presupposes three components: domain-relevant skills, creativity-relevant processes (such as divergent thinking) and intrinsic motivation. Some psychologists include factors of the environment. Sternberg (2003) has suggested the following five components of creativity: (1) expertise: a well-developed base of knowledge; (2) imaginative thinking skills: the ability to see things in novel ways, to recognise patterns, to make connections; (3) a venturesome personality: to seek new experiences, tolerate ambiguity and risk, and persevere in overcoming obstacles; (4) intrinsic motivation: to be driven more by interest, satisfaction and challenge than by external factors; and (5) creative environment: sparking, supporting and refining ideas (see Jackson et al. 2012).

As philosophers, we are not so much interested in measuring or testing creativity, but rather in understanding the mental states, processes, concepts etc. involved in creativity. Contemporary creativity research in philosophy is still in its infancy (as Elliot S. Paul and Scott B. Kaufman put it: '"the philosophy of creativity" is still a neologism in most quarters', Paul and Kaufman 2014, 4), and many authors have followed the main idea (probably taken from psychology) that creativity should be seen as

the feature of a subject, more precisely as an ability: Boden's definition of creativity as the ability to come up with ideas or artefacts that are new, surprising, and valuable is, for instance, followed by Sternberg and Lubart (1999, 3), according to which 'creativity is the ability to produce work that is both novel (i.e., original, unexpected) and appropriate (i.e., useful, adaptive concerning task constraints)'. Similarly, Hills and Bird (2019, 1) state explicitly: 'We focus here on the definition of creativity as a trait of persons'.

Let me compare the case of creativity with the case of empathy. In an attempt to understand empathy, psychologists design experiments in which they measure the subjects' degree of empathy by conducting experiments designed to trigger empathy (e.g., Keith Oatley's work on the effect of reading literary fiction on empathy, see Oatley 2012; 2016). Sometimes, the notion of empathy includes helping behaviour. Now suppose empathy was defined as the ability to react appropriately to a person in need, i.e. display helping behaviour, where this involves understanding the other person from their own perspective. Such an account of empathy would fail to do justice to the following two facts: first, that empathy is a mental state (or process) with certain phenomenal and structural properties and second, that there could be other ways of reacting appropriately which are indistinguishable from appropriate empathic reactions, for instance reactions involving sympathy (feeling for, but not with the other person). As philosophers, we are interested in the mental state or states on the one hand, and in the relation between the mental states and helping or moral behaviour on the other hand, and hence we make sure to distinguish empathy from the appropriate behaviour in response to a person in need. Our research concerns both separately: structural and phenomenological aspects of the mental states involved in empathy, and normative questions concerning the relation between empathy and helping or moral behaviour.

Similarly, we should distinguish creative processes from creative outputs. When studying creativity, we should not primarily look at the creative person as a person who is able to produce a certain kind of output. Sometimes, a creative process does not lead to a creative output, and sometimes we have creative outputs that do not involve a creative process. Instead of looking at creativity as an ability to produce a certain output, we should look at it as a mental process on the one hand, and as a feature of certain products on the other hand, and we should keep both apart.

Looking at product and process separately helps us realise what is wrong with Boden's distinctions. The distinction between two levels, psychological and historical creativity, according to which psychological creativity covers cases in which something is novel for the creative subject only, allows us to cover the fact that we can be creative no matter how new the product is. Cases of psychological creativity are mostly

cases in which we are interested in the creative process. Historical creativity, in contrast, concerns a feature of the creative product. We saw that the distinction between three types of creativity ('combinational', 'exploratory' and 'transformational' creativity) is independent of our reaction to the product, and it is also independent of the mental process that leads to the product. Hence, the three types of creativity should be attributed solely to the creative product.

Of course, just like empathy often leads to helping behaviour, a creative process often leads to a creative product. But this does not mean they should not be looked at and accounted for separately.

Process creativity and product creativity

When it comes to creativity, we care about two different, though often related things: the process of creating and the creative product. Keeping them apart can resolve tensions between seemingly conflicting intuitions. I will first address the case of creativity and value, because it helps us understand some differences in both notions. Second, I will discuss AI creativity, and third, I will show that two notions of creativity also help us to think about animal creativity.

Ad c) Creativity and value. We have seen that value is usually thought to be the aspect that distinguishes creativity from mere originality. However, Alison Hills and Alexander Bird (2019) have recently argued that value is not necessary for creativity. Following psychological approaches, Hills and Bird focus on creativity as a feature of a subject. They define creativity as the set of dispositions of a person to have novel ideas (originality), generate those ideas through the use of imagination (imagination), produce many such ideas (fertility), and carry through these ideas to completion (motivation) (Hills and Bird 2019, 2). Fertility—in psychology 'fluency'—and originality are, besides flexibility and elaboration, classic components of the original *Torrance Test of Creative Thinking*.

If value is not part of the definition of creativity, some creative ideas are only original, but not good or useful, etc. Nikola Tesla and William Herschel (the latter discovered infrared radiation), for instance, did not only have valuable or true ideas; some of their ideas were bad and false. According to Hills and Bird, 'it's simply not plausible that one set of dispositions (creativity) produced the good ideas of Tesla and Herschel and quite another generated the bad ones' (Hills and Bird 2019, 12). We can call these bad ideas a 'manifestation of creativity, at least of a minimal kind', because they are the result of a creative, though kind of unsuccessful process.

It is worth taking note of the fact that this account emphasises the role of the imagination in creating. Similarly, Arcangeli (forthcoming) argues that imagination is necessary for creativity. When engaged in the

imaginative process of creating, other mental states, in particular evaluative mental states, play a role as well: We are usually aiming to create either something aesthetically valuable, or something useful, or appropriate, etc., depending on the context and the kind of idea or artefact we are aiming at. For these and probably other reasons we usually value creative processes.

Similar to Boden's definition, Hills and Bird's approach aims to comprise both the mental process of creating and the creative product in one definition of creativity. Hills and Bird realise that the process of creating is crucial, but they call it 'deeply misguided' to approve of creativity even when the output is not valuable (2019, 18). Keeping product creativity and process creativity apart, it can be acknowledged that we can approve of the creativity of a process without calling the output valuable. We can simply say that Tesla's and Herschel's creative processes are valuable, but only some of the products are valuable and hence deserve to be called 'creative'. Hills and Bird also note that

> not all ways of producing novel, valuable ideas are creative. For instance, it may be possible to produce such ideas by a random process; or by a purely mechanical process, such as following a simple rule. But these are not typically exercises of creativity.
>
> (Hills and Bird 2019, 2)

The authors correctly note that the processes involved in the generation of a creative product are sometimes not creative processes. But keeping product creativity and process creativity apart, we can say that the product can be creative, while the process is not (I will discuss an example in the section 'Monkey selfies and cross-cultural differences').

Further, Hills and Bird focus on the creative person: In order to be a creative subject, an individual has to go through many creative processes, and has to actually produce creative outputs. My suggestion is to care about the creative subject only when we are concerned with psychological matters such as why people are creative, how we can train people to be more creative, which other features in humans correlate with creativity, etc. These are, of course, interesting and important questions. But when we are interested in the nature of creativity, we should be interested in the mental process (just as we are interested in empathy as a mental state or process) on the one hand, and in the conditions under which we call a certain product 'creative' (just as we are interested in helping or moral behaviour). The discussion of the mental process involves connections and involvement of evaluative states and other mental states, and the discussion of what counts as a creative product involves discussion of different kinds of value, different levels of novelty (historical novelty or novelty relative more narrow contexts), and different kinds of surprise (Boden's distinction of three types of creativity).

If we quit attributing creativity primarily to subjects, we can see that in cases that apparently do not involve value, what is actually creative and should be called creative is the mental process, but not the product. This creative process very likely comes with subjective value, which explains why creative activities are cherished and recommended. It also explains why we value a child's drawing: not because the drawing is valuable (we would not be able to sell it, for instance), but because it is an expression of a valuable creative process.

Ad b) AI creativity. Most authors accept that novelty (and surprise) and value are necessary but insufficient for a definition of creativity, but they disagree about what else is needed, e.g., a certain phenomenology (Nanay 2014), flair (Gaut 2003), intentionality (e.g. Boden 2014), agency (Paul and Stokes 2018), or imagination (Hills and Bird 2019). The underlying concern is to account for the idea that the creative output has to be produced in the right kind of way: Creativity cannot be the result of a mechanical or imitative process. This of course makes it questionable whether computer processes can be creative.

But if we keep product creativity and process creativity apart, we can ascribe creativity to products generated by AI as long as they are new, surprising and valuable. We can moreover explain a case that has been raised by Matthew Kieran. Berys Gaut presents it as follows:

> Suppose that you daub me all over with paint and imprison me in a dark room in which there is a primed canvas. I fall around for several hours, attempting to escape; my frantic thrashings cover the canvas in such a way that it becomes, unknown to me, a stunningly good abstract painting, significantly different in appearance from any abstract painting hitherto produced. I have inadvertently produced something valuable and original, but it would be wrong to say that I have done so creatively—I made it purely by chance. Or suppose that I engage in a mechanical search procedure form some desired outcome, systematically working through all the relevant possibilities, and in the course of the search come across a result that is original and valuable. Again, the upshot of such a search procedure is not an instance of creativity, for the procedure adopted is a mechanical one.
>
> (Gaut 2003, 270)

According to Gaut, the making of a creative product must involve what he calls 'flair'. Gaut further claims that the same definition holds for the process and the product:

> A process is creative when it is the producing of something valuable and original by flair (or, if we allow that a creative process need not always produce a creative outcome, when it is an instance of the

> kind of process involving flair that usually tends to produce original and valuable things). And artefacts (in a broad sense including the performance of acts) are creative when they are original, valuable and produced by flair.
>
> (Gaut 2003, 271)

We can see that Gaut is willing to call the creative process 'creative' even if it does not generate a valuable product. We could thus also be willing to judge the product independently. If the painting is indeed valuable and original, we can call it creative. However, in some cases of abstract or conceptual art, what we value may be the creative process, and we may value the painting in a derivative sense, just like we value the drawing of a child because it is an expression of a creative process. However, it seems that if we care enough about the output, we often do not and arguably should not care about the process.

Note also that if we keep product creativity and process creativity apart, AI creativity cannot be a threat to human creativity, because it only concerns the product, and process creativity will be valuable independently.

Ad a) Animal creativity. We have seen that distinguishing between different types and levels of creativity allows us to account for animal creativity. Until recently, a majority of creativity researchers had restricted their research to human creativity (however, see Kaufman and Kaufman 2014). Ronald A. Beghetto writes:

> Why do creativity researchers, such as myself, fail to acknowledge nonhuman creativity? One reason I became aware [. . .] had to do with the way we categorize our beliefs about creativity. The issue for me (and I suspect other creativity researchers) boiled down to viewing nonhuman creativity as a difference in kind rather than a difference in degree.... Once we endorse this view, it becomes easy to privilege human creative behavior.
>
> (Beghetto 2015, 27)

Beghetto thinks that the idea that human creativity is much different from animal creativity rests on the idea that human creativity is a kind of intentional behaviour (deliberate behaviour) and involves a kind of decision. Once we realise that human and animal creativity differ only in degree and not in kind, we can admit that there are different levels of intentions and different levels of creativity (Beghetto 2015, 28).

But what do we care about when it comes to animal creativity? Some of the creative products generated by animals may be interesting and creative in their context, but researchers are mostly interested in whether animals are similar to us or in what exactly distinguishes us from them, that is, the creative process. There are, however, cases in which animals

have generated new, surprising, and valuable products. I will discuss one such case in the next section.

Monkey selfies and cross-cultural differences

In this section, I will give two independent kinds of evidence in favour of distinguishing two different notions of creativity. The first shows that we are willing to separate the notion of creativity from certain mental processes in humans, and the second supports the idea of two different notions by pointing out that cross-cultural studies reveal two different aspects of creativity.

In his paper 'I think, therefore I invent: creative computers and the future of patent law' (2016), Ryan Abbott argues that computers already generating creative output should be considered inventors under the Patent and Copyright Clause of the US Constitution. He claims that current US patent law does not account for the fact that computers meet the requirements to be understood as inventors: 'There is no statute addressing computational invention, no case law directly on the subject, and no pertinent Patent Office policy' (Abbott 2016, 1080). Abbott thinks that granting computers inventorship would further the development of creative computers and thereby lead to new inventions. While his arguments are driven by practical concerns, Abbott presents some insight into the development of patent law in the US, in particular into the trend away from the requirement of a particular mental act of creating.

One factor that contributed to current patent law was a case of the Monkey Selfies. The Monkey Selfies are a series of photographs taken in 2011 with a camera owned and set up by nature photographer David Slater. A Celebes crested macaque took the pictures with a remote trigger. David Slater licenced the photographs and claimed copyright, but Wikimedia Commons and other parties reposted them, claiming that they belong to the public, since Slater did not take them himself.

In response to the following legal disputes between Slater and Wikimedia Commons, the Copyright Office published an updated 'Human Authorship Requirement' in its 2014 compendium, which states:

> To qualify as a work of "authorship" a work must be created by a human being [. . .] . The Office will not register works produced by nature, animals, or plants [. . .] . Similarly, the Office will not register works produced by a machine or mere mechanical process that operates randomly or automatically without any creative input or intervention from a human author.
>
> (Quoted from Abbott 2016, 1099–1100)

In 2015, People for the Ethical Treatment of Animals (PETA) filed a copyright suit against Slater and his publishing company, using the

so-called next-friend principle, which allows one to file a law suit on behalf of a person unable to do so. PETA claimed that Naruto, the monkey who had taken the selfies, should own the copyrights of the Monkey Selfies. While the court eventually decided that a monkey cannot legally own copyrights, PETA and Slater settled on an agreement according to which Slater would donate a certain percentage of the money he would make with the Monkey Selfies to wildlife organisations.

Long before this dispute, the notion of creativity and what it involves had already been discussed in the context of US patent law. In 1952, the American congress abolished the so-called 'Flash of Genius doctrine', which held that 'in order to be patentable, a new device, "however useful it may be, must reveal the flash of creative genius, not merely the skill of the calling"' (quoted from Abbott 2016, 1108). The doctrine meant that an invention must have occurred through a specific mental process, namely a 'flash of genius' rather than as a 'result of long toil and experimentation' (Abbott 2016, 1108). Interestingly, the 'Flash of Genius doctrine' was only in place for about a decade. Because it involved making a judgement about a person's mental states or processes, it was simply too difficult to implement.

Abbott claims that since patent law is already indifferent concerning the exact nature of the mental state involved in human creating, it should generally be indifferent concerning how the invention came about, and hence it should acknowledge computer inventorship: 'patentability of computational inventions should be based on the inventiveness of a computer's output rather than on a clumsy anthropomorphism because [. . .] patent law should be interested in a functionalist solution' (Abbott 2016, 1111). However, current US patent law still requires that inventions be the result of a mental act, and hence 'it is unclear whether a computer autonomously conceiving of a patentable invention could legally be an inventor' (Abbott 2016, 1097). Abbott argues that it is not clear how this requirement ought to be understood, and whether computers that generate products that are, in principle, patentable would have to mimic human mental processes, or whether they would only have to engage in their very own kind of creative process (Abbott 2016, 1108).

The copyright dispute concerning the Monkey Selfies suggests that we (or some of us) are, in fact, willing to separate the creative product from the creative subject undergoing certain mental states. The positions of both of Slater's opponents presuppose that the notion of the creative product can be detached from the requirement of a human creator. While PETA thinks the animal should be recognised as the creator, Wikimedia operates under the assumption that nobody fills this position. In both cases, the assumption is that there is no human mental process of creating involved. A separation of the creative product from the question of how it came about is exactly what Abbott demands for creative outputs by AI. The Monkey Selfies thus present a case in which an animal has produced a valuable product, and they also show that we are willing to look at this product as a creative product.

Here is the second piece of evidence in favour of two notions of creativity. In their study published in a paper titled 'In the World or in the Head: External and Internal Implicit Theories of Creativity' (2011), Susannah Paletz, Kaiping Peng, and Siyu Li investigate the concept of creativity in different cultures, mainly in East Asian and Caucasian North American individuals. They build their study on the assumption of different styles of attribution across cultures (going back to the famous individualism-collectivism framework, see Hofstede 1980). For instance, while North Americans are more likely to ascribe agency to individuals, East Asians are more likely to ascribe agency to a collective (Menon et al. 1999), and East Asians are more likely to take contextual aspects into account than North Americans, who are focused on internal and individual aspects (Choi et al. 2003; see Paletz et al. 2011, 84). Applying this assumption about different styles of attribution to creativity, the authors predicted that to a person with their main focus on internal aspects, 'creativity is more likely to be considered to encompass creative activities such as reflection, states of being and inner experience, intuition, thinking, and high levels of self-awareness', whereas a person who is more aware of situational aspects, 'might focus on more external dimensions in his or her conception of creativity' (Paletz et al. 2011, 85). Such expressions of creativity would be, amongst other things, visible products. They further predicted that East Asians are significantly more likely to emphasise external factors in creativity, and North Americans are more likely to emphasise internal factors (Paletz et al. 2011, 86).

The authors were able to confirm their hypotheses and conclude that this does not mean that North Americans 'disdain accomplishments and external evidence for creativity', or that East Asians 'see all internally-focused activities as noncreative' (Paletz et al. 2011, 98). Their study is directed against previous studies in which, drawing from anthropological, philosophical, and psychological literature, it was suggested that the Eastern notion of creativity emphasises 'inner processes and fulfilment', while the Western notion of creativity emphasises creative products (Lubart 1999).

For our purposes, it does not matter which culture focuses on which aspects of creativity, and a much weaker conclusion is sufficient: There are two different ways of attributing creativity, of which one places importance on external aspects such as the creative product and the context, and the other pays attention to the creative individual's mental states involved in the process of creating. I take this to be evidence for the claim that we should distinguish two different notions of creativity.

Conclusion

In this chapter, I have argued that distinguishing two notions of creativity can resolve apparent conceptual tensions and can carve the way

for future philosophical research. I have not given an account of either notion of creativity, but creativity research in philosophy should address creativity as a valuable, imaginative process on the one hand, and it should address the question of when a product is creative within a certain context on the other hand. If we make this distinction, we do not need Boden's two levels of creativity, psychological and historical creativity, and we can explain better how animals, AI generated art and inventions, and exceptional humans on the one hand and children's drawings on the other hand can all be called 'creative'.

Notes

1 While computers can also learn in ways similar to humans, they cannot be curious; see Glock and Nickl (2020).
2 See https://www.kunstaufraeumen.ch. Last accessed: 30 March 2022.

References

Abbott, R. (2016). I Think, Therefore I Invent: Creative Computers and the Future of Patent Law. *Boston College Law Review* 57(4): 1079–1126.

Amabile, T.M. (1988). A Model of Creativity and Innovation in Organizations. In Staw, B.M. and Cummings, L.L. (eds.), *Research in Organizational Behaviour* 10. Greenwich, CT: JAI Press, pp. 123–167.

Arcangeli, M. (2022). The Creative Side of Recreative Imagination. In Engisch, P. and Langkau, J. (eds.), *The Philosophy of Fiction: Imagination and Cognition*. Routledge, pp. 77–96.

Beghetto, R.A. (2015). Commentary on Chapter 1: What Can Creativity Researchers Learn from Grey Parrots? In Kaufman, A.B. and Kaufman, J.C. (eds.), *Animal Creativity and Innovation*. London Wall: Elsevier Academic Press, pp. 25–29.

Boden, M.A. (1994/2004). *The Creative Mind: Myths and Mechanisms*. New York: Routledge.

Boden, M.A. (2010). *Creativity and Art. Three Roads to Surprise*. New York: Oxford University Press.

Boden, M.A. (2014). Creativity and Artificial Intelligence: A Contradiction in Terms? In Paul, E.S. and Kaufman, S.B. (eds.), *The Philosophy of Creativity. New Essays*. New York: Oxford University Press, pp. 224–244.

Choi, I., Dalal, R., Kim-Prieto, C. and Park, H. (2003). Culture and Judgment of Causal Relevance. *Journal of Personality and Social Psychology* 84: 46–59.

Gangadharbatla, H. (2021). The Role of AI Attribution Knowledge in the Evaluation of Artwork. *Empirical Studies of the Arts* 40: 125–142. https://doi.org/10.1177/0276237421994697

Gaut, B. (2003). Creativity and Imagination. In Gaut, B. and Livingston, P. (eds.), *The Creation of Art: New Essays in Philosophical Aesthetics*. Cambridge: Cambridge University Press, pp.148–173.

Gaut, B. and Kieran, M. (2018). *Creativity and Philosophy*. New York: Routledge.

Glock, H.-J. (2019). Agency, Intelligence and Reasons in Animals. *Philosophy* 94(4): 645–671.

Glock, H.-J. and Nickl, R. (2020). Computer sind nicht neugierig. *UZH Magazin: die Zeitschrift der Universität Zürich: die Wissenschaftszeitschrift* 1: 35–37.

Guilford, J.P. (1956). The Structure of Intellect. *Psychological Bulletin* 53(4): 267–293.

Hennessey, B.A. and Amabile, T.M. (2010). Creativity. *Annual Review of Psychology* 61: 569–598.

Hills, A. and Bird, A. (2019). Against Creativity. *Philosophy and Phenomenological Research* 99(3): 694–713.

Hofstede, G. (1980). *Culture's Consequences: International Differences in Work Related Values*. Beverly Hills, CA/London: Sage.

Jackson, L.A. et al. (2012). Information Technology Use and Creativity: Findings from the Children and Technology Project. *Computers in Human Behavior* 28(2): 370–376.

Kant, I. (1790/2000). *Critique of the Power of Judgment*. Edited by Paul Guyer. Translated by P. Guyer, and E. Matthews. Cambridge: Cambridge University Press.

Kaufman, J.C. and Beghetto, R.A. (2009). Beyond Big and Little: The Four c Model of Creativity. *Review of General Psychology* 13(1): 1–12.

Kaufman, A.B. and Kaufman, J.C. (2014). Applying Theoretical Models on Human Creativity to Animal Studies. *Animal Behavior and Cognition* 1: 78–90. http://doi.org/10.12966/abc.02.01.2014

Lubart, T.I. (1999). Creativity across Cultures. In Sternberg, R.J. (ed.), *Handbook of Creativity*. New York: Cambridge University Press, pp. 339–350.

Manley, J. (2020). The Jewel in the Corona: Crisis, the Creativity of Social Dreaming, and Climate Change. *Journal of Social Work Practice. Psychotherapeutic Approaches in Health, Welfare and the Community* 34(4): 429–443.

Menon, T., Morris, M.W., Chiu, C.-Y., and Hong, Y.-Y. (1999). Culture and Construal of Agency: Attribution to Individual versus Group Dispositions. *Journal of Personality and Social Psychology* 76: 701–717.

Nanay, B. (2014). An Experiential Account of Creativity. In Paul, E.S. and Kaufman, S.B. (eds.), *The Philosophy of Creativity*. New York: Oxford University Press, pp. 17–23.

Novitz, D. (1999). Creativity and Constraint. *Australasian Journal of Philosophy* 77: 67–82.

Oatley, K. (2012). The Cognitive Science of Fiction. *WIREs Cognitive Science* 3: 425–430.

Oatley, K. (2016). Fiction: Simulation of Social Worlds. *Trends in Cognitive Sciences* 20(8): 618–628.

Paletz, S.B.F., Peng, K. and Li, S. (2011). In the World or in the Head: External and Internal Implicit Theories of Creativity. *Creativity Research Journal* 23(-2): 83–98.

Paul, E.S. and Kaufman, S.B. (2014) (eds.). *The Philosophy of Creativity. New Essays*. New York: Oxford University Press.

Paul, E.S. and Stokes, D. (2018). Attributing Creativity. In Gaut, B. and Kieran, M. (eds.), *Creativity and Philosophy*. New York: Routledge, pp. 193–209.

Pepperberg, I.M. (2015). Chapter 1- Creativity and Innovation in the Grey Parrot (Psittacus erithacus). In Kaufman, A.B. and Kaufman, J.C. (eds.), *Animal Creativity and Innovation. Explorations in Creativity Research*. New York: Elsevier Inc., pp. 3–29.

Sawyer, R.K. (2012). *Explaining Creativity: The Science of Human Innovation*. New York: Oxford University Press.

Sternberg, R.J. (2003). Creative Thinking in the Classroom. *Scandinavian Journal of Educational Research* 47: 325–338.

Sternberg, R.J. (2006). The Nature of Creativity. *Creativity Research Journal* 18(1): 87–98.

Sternberg, R.J. and Lubart, T.I. (1999). The Concept of Creativity: Prospects and Paradigms. In Sternberg, R.J. (ed.), *Handbook of Creativity*. Cambridge: Cambridge University Press, pp. 3–15.

Stokes, D. (2011). Minimally Creative Thought. *Metaphilosophy* 42: 658–681.

Stokes, D. (2016). Imagination and Creativity. In Kind, A. (ed.), *The Routledge Handbook of Philosophy of Imagination*. New York: Routledge, pp. 247–261.

Tomasello, M. (2000). Two Hypotheses about Primate Cognition. In Heyes, C. and Huber, L. (eds.), *The Evolution of Cognition*. Cambridge, MA: MIT Press, pp. 165–183.

Part IV

Normativity and reasons

15 Rationality, reasons, and rules

Brad Hooker

Introduction

I had already started learning from Hanjo Glock before I joined him at Reading in 1993. Once I became his colleague, the rate at which I absorbed from him philosophical arguments and insights and wider wisdom was humbling. And the example he set—of ambition, energy, conscientiousness, fortitude, wit, and good humour—inspired everyone around him. I was in awe of him then, and this awe has increased in the years since, as his arguments have gained adherents across many areas of philosophy.

Rationality

Glock writes,

> Within contemporary academic debates one can distinguish four general conceptions of rationality. According to the first, it is the capacity to maximise satisfaction of one's interests or goals; according to the second, it is responsiveness to reasons; according to the third, it is the ability to reason—draw theoretical and/or practical inferences and to avoid inconsistencies; according to the fourth, it is the ability to justify one's actions and beliefs to others.
>
> (2019a, 665)

I agree that these four general conceptions of rationality are especially prominent. And I think Glock and I agree which of these four are most plausible.

Before I start on explaining why, let me make a point of clarification about what I think Glock meant by 'conceptions of rationality'. I think he meant competing conceptions of what constitutes rationality. *Having rationality* is compatible with *being irrational*. Having rationality is a matter of possessing capacities, and being rational is a matter of correctly exercising those capacities.

DOI: 10.4324/9781003202929-20

The first conception of rationality that Glock lists is the idea that rationality is the capacity to maximise satisfaction of one's interests or goals. This idea is not as simple as it initially looks. Admittedly, there are simple cases. For example, suppose your chief desire right now is for a cold drink and so you go now to the refrigerator to take water from the pitcher there. In other cases, however, you have a desire that needs *further specification* before you start thinking about means to satisfy it. For example, you want to achieve something important this week, but you need to think about whether finding a merely temporary solution to a problem afflicting your community qualifies as an important achievement (cf. Williams 1981, 104). Another example might be that you want to find something that appropriately symbolises your friendship with Kayta, but you have to think about whether a beautiful bowl or a long-life house plant would be better.

The idea that rationality consists in the *capacity* to maximise satisfaction of one's interests or goals is often associated with the idea that rationality *requires* one to exercise this capacity if one has the capacity. Suppose my *sole* goal right now is to get home by the quickest route and, fully aware of what I am doing, I turn right, despite knowing that turning right will not get me home by the quickest route. The previous sentence might strike you as somehow inaccurate. You might surmise, for example, that getting home by the quickest route must not have been my only goal, or that I must have failed to appreciate that turning right would take me away from the quickest route. However, if you accepted that I really did have only one goal and that I really did appreciate that turning right would not achieve that goal, then you would think that I was irrational, indeed that what I did was barely intelligible, given my mental states at the time. How could I really have getting home by the shortest route as my goal, know that turning right would not achieve that goal, and yet turn right in order to achieve the goal?

Actually, people sometimes do have a goal, know that a certain kind of action will not achieve the goal, and yet choose an action of that kind in order to achieve the goal. For example, I want to impress you, I know that bragging will not impress you, and yet I find myself bragging in order to impress you. Sometimes I behave irrationally. Sometimes other people do too.

In many situations, one does not know in advance what all the consequences of this or that act would turn out to be. Thus, in many cases, one does not know in advance what the most efficient means to one's goal would be. In such cases, rationality cannot reasonably insist that one choose the most efficient means, since one does not know which means would be most efficient.

In cases of uncertainty, one often has *some* information about the *probabilities* of outcomes of possible actions. If the information I have suggests that turning left will very probably get me home sooner than

not turning left, then, as long as my sole aim is to get home as quickly as possible, rationality requires me to turn left. More generally, when facing a choice between actions whose *actual* consequences one cannot know in advance, one would be irrational to ignore evidence about *likely* consequences (unless one has grounds to think this evidence is misleading). A way of incorporating probabilities into one's account of rationality is to conceive of rationality as combining judgements about the values of achieving one's different goals and the probabilities that alternative actions would achieve those goals.

Now we should turn to the question of whether rationality requires one to prioritise one's own interests and goals over other people's. The idea that rationality consists in the capacity to maximise satisfaction of one's interests or goals does *not* dictate that one's goals must be focused on one's own life. Whether your goal is to support Amnesty International or the United Nations International Children's Emergency Fund or your impoverished cousin, rationality conceived of as the capacity to maximise satisfaction of one's interests or goals does not denigrate your altruistic goals. Indeed, this conception of rationality is completely non-judgemental about your goals and indeed about everyone else's.

While this conception of rationality does not denigrate altruistic goals, it also does not insist one have altruistic goals or concerns. My own view is that there is something rationally defective about agents who are concerned exclusively with their own goals and attach no non-instrumental importance to other people's worthwhile goals.

But a reply to my view might be that the concept of individual agency draws a sharp line between the agent's goals and intentions and other people's goals and intentions, and that thus *rational* individual agency cannot discard this sharp line. I agree that the concept of individual agency would be severely threatened if each agent were rationally required to give the goals of every other agent the same weight as his or her own. But my much more modest contention is that rational agents must attach *some* non-instrumental importance to other people's worthwhile goals. This modest contention hardly threatens the concept of agency.

I admit that taking the distinction between one's own goals and those of others to be practically important is not arbitrary, even if this distinction can be taken too far. However, there are other possible distinctions that are, beyond question, arbitrary. One is Derek Parfit's example of someone who cares about benefits or harms to himself that occur on any day of the week except Tuesday (Parfit 1984, 124–126). Here, someone singles out an arbitrary period of time—Tuesdays. Someone else might single out an arbitrary space. She might, for example, care about what benefits or harms herself anywhere except in Winnipeg.

Both the time and space kind of examples have an infinite number of instances. Just as it is irrational to care about what happens to oneself *except on* Tuesday, it is irrational to care about what happens to oneself

except on Saturday, or *except at* two minutes to the hour, etc. Likewise, it is irrational to care about what happens to oneself *only on* the 15th of the month. And just as it is irrational to care about what happens to oneself *except in* Winnipeg, it is irrational to care about what happens to oneself *except in* some other particular place, or *only in* some other particular place.

In response to these kinds of examples, we might point out that the people in these examples do not deny that they have interests and goals on Tuesdays, just as they have interests and goals on other days of the week, and at two minutes to the hour, and at any other particular time, and when they are in Winnipeg. Hence, if rationality requires one to maximise the satisfaction of one's interests and goals, then the satisfaction of one's on-Tuesday interests and goals and one's in-Winnipeg interests and goals *are* to be counted in the aggregate of one's interests and goals to be maximised. And if rationality can require one to care about, or at least count in one's calculation of what is to be maximised, one's future interests and goals even though one does not *now* care about them, it is not such a big step to holding that rationality can require one to give some weight to the interests and goals of others even though one does not *now* care about them.

I turn now to the second conception of rationality that Glock lists, the conception of rationality as responsiveness to reasons. If we think that people *have reason* to maximise the satisfaction of their interests and goals, then we may well think that *rationality* calls for people to *maximise* satisfaction of their interests and goals. One of the most often discussed examples of irrationality is the pursuit of present benefits when these are known to lead to greater harms later (Sidgwick 1907, 124, n. 1, 381; Nagel 1970, chs. 6–8; Parfit 1984, 133, 158–163). Imagine someone who gives himself a weekend of leisure though he knows he needs to be studying assiduously throughout the weekend in order to pass the test he faces next week. It is very tempting to say that he is irrational in choosing the relatively minor present benefit of a weekend of leisure over the far more significant future benefits that would come from passing the test.

The benefits to him that would come from passing the test provide reasons for him to do what is necessary and sufficient for him to pass the test. He is much more likely to pass the test if he tries to pass it, and if he studies assiduously throughout this weekend for the test. Admittedly, the pleasure he would get from a weekend of leisure would constitute a benefit to him, and the prospect of this benefit generates a reason for devoting the weekend to leisure. However, the weights of reasons deriving from benefits or harms to him presumably correlate with the sizes of the benefits or harms to him, and the benefits to him of passing the test would be much greater than the benefit to him of a pleasurable weekend. Hence, he presumably has stronger reason to spend the weekend preparing for the test than he does to spend the weekend on leisure.

If we conceptualise rationality as responsiveness to reasons, we might say that what is irrational about indifference to benefits or harms to oneself that occur on Tuesday or in Winnipeg is *not* that such indifference *arbitrarily* distinguishes between benefits or harms to oneself on Tuesday and benefits or harms to oneself on other days, or between benefits or harms to oneself when one is in Winnipeg and benefits or harms to oneself when one is in other places. What is irrational about indifference to benefits or harms to oneself that occur on Tuesday or in Winnipeg is instead that such indifference is unresponsive to one's reasons to care about benefits or harms to oneself that occur on Tuesday or in Winnipeg.

Putting off a more elaborate discussion of reasons until the following section, I turn now to the third conception of rationality in Glock's list, rationality as drawing theoretical and/or practical inferences and avoiding inconsistencies. I think Glock and I agree that this conception of rationality is the most widely shared. Having obviously inconsistent beliefs is irrational. Perhaps failures to draw obvious inferences is also irrational. But we go too far if we insist that all failures to draw unobvious inferences are irrational.

Drawing valid inferences can come at a cost—in terms of effort, time, and cognitive or emotional overload. Because of such costs, rationality must concede that drawing further valid inferences can sometimes not be what agents have most practical reason to do. Imagine some agent sitting in an ivory tower teasing out valid inferences while the city around her burns.

I turn now to the fourth conception of rationality in Glock's list, rationality as the ability to justify one's actions and beliefs to others. In the case of each of the first three conceptions of rationality Glock listed—the capacity to maximise satisfaction of one's interests or goals, responsiveness to reasons, and the ability to draw theoretical and/or practical inferences and to avoid inconsistencies—I have mentioned these as capacities but mostly discussed them as requirements, in the form of 'rationality is not only the capacity to [. . .] but also the requirement to [. . .]. '. In the case of the fourth conception in the list, there seems to me no pressure to think of rationality as a requirement as well as a capacity. Admittedly, in many contexts, there are requirements to justify one's actions and beliefs to others. However, these requirements do not come from rationality.

Justifying one's actions and beliefs to others must be understood as laying out considerations and reasoning which others *might* accept as justifying one's actions and beliefs. Justifying one's actions and beliefs to others *need not entail* that all others would *in fact* accept one's justifications. Some people might reject these justifications because they cannot see why one's conclusions follow from one's premises. And some people might deny one or more of one's premises.

Laying out the considerations and reasoning that justify one's actions and beliefs requires thought—and, often, fairly sophisticated thought.

We need not here get caught up in the question of how sophisticated such thought must be. If rationality consists in the ability to justify one's actions and beliefs to others, one implication is that, whatever the level of sophistication that thought must have in order to lay out for others the considerations and reasoning justifying one's actions and beliefs, any being who is incapable of that level of sophisticated thought is incapable of rationality. And it does seem intuitively correct that any being incapable of a threshold level of sophisticated thought lacks rationality.

What does *not* seem intuitively correct is that you cannot be rational unless you have both the capacity and the opportunity to justify *to others* your beliefs and actions. On the face of things, your ability to *lay out* for others your considerations and reasoning requires that you have the ability to *communicate* these considerations and lines of reasoning. Yet your having the ability *to communicate* is not a necessary condition of your being rational, much less of your having rationality. Imagine that you survived a terrible injury that robbed you of the capacity to communicate your mental states. Obviously, your *loss of the ability to communicate* your mental states does not entail that you *lack mental states* or that you *lack abilities to think, draw valid inferences, entertain hypotheses, evaluate arguments, assess actual and possible beliefs and actions, form desires and intentions, etc*. If you have retained these abilities, then you have rational capacities. You are able to think, desire, and intend rationally, even if you cannot communicate to others what you think, intend, etc.

Perhaps the most charitable way to construe the idea that rationality is the capacity to justify one's beliefs and actions to others circumvents problems about the inability to communicate with others, whether that inability comes from one's injuries or from the absence or deafness of others. Rationality as the capacity to justify one's beliefs and actions to others should be construed as the capacity to lay out, at least in one's own mind, the considerations and reasoning that one takes to support one's beliefs and actions. Possible others might accept them if presented with them.

Reasons

I indicated that I would return to the conception of rationality as responsiveness to reasons (an old haunt; see Hooker 1987). This conception of rationality seems to me one of the two most plausible conceptions of rationality, the other being the conception of rationality as drawing inferences and avoiding inconsistencies. Actually, rationality as drawing inferences and avoiding inconsistencies could be thought of as a subset of rationality as responsiveness to reasons (cf. Kiesewetter 2017; Lord 2018). To draw inferences and to avoid inconsistencies are ways of responding to reasons. For example, you have conclusive reason not to believe more than one of any two inconsistent propositions.

Inferences from premises about probabilities and practical inferences are more complicated. Suppose I knew the homemade curry put in front of me was 50% likely to upset my digestion. Suppose I knew both that I was not hungry and that there was another food choice available to me that was equally delicious and nutritious but very unlikely to upset my digestion. Still, maybe I had most reason to eat the curry because I would have offended my boss if I had not (he was the one who made the curry).

This example illustrates an important aspect of reasons for desiring, intending, and acting. That the curry had a 50% likelihood of making me feel unwell was a reason not to eat it. There might have been other reasons not to eat it, such as that, while eating the curry, I would have been likely to spill it on myself. And yet all the reasons against eating the curry might be outweighed by reasons to eat it. (With respect to reasons for belief, of course there can be evidence *for* some conclusion and evidence *against* that same conclusion.)

One of the great breakthroughs in practical philosophy was W. D. Ross's distinction between prima facie duties and all-things-considered duties (Ross 1930, ch. 2). What Ross meant by 'prima facie duty' was not 'duty on first look', as if the duty would turn out not to be a duty on subsequent investigation. What he meant was 'duty in one respect and to an extent'. Many philosophers have thus changed Ross's terminology from 'prima facie' to 'pro tanto', in order to better express what Ross had in mind.

Ross proposed that, when we face a choice between mutually exclusive acts, we are to weigh up the different pro tanto duties we have in the situation and decide which act has the strongest aggregate of pro tanto duty on its side. The act with the strongest aggregate of pro tanto duty on its side is then the act that is our all-things-considered duty in this situation. The idea that there is a plurality of moral pressures, which do not come in a strict hierarchy of importance, and that moral agents thus need good judgement to adjudicate conflicts between these moral pressures was the central idea of Ross's deontological pluralism in the 1930s, though he was hardly the only proponent. And moral pluralism is best expressed using the pro tanto/all-things-considered distinction.

For example, you have a pro tanto duty to help those in need, especially when you can do so at no cost to yourself. Suppose I am in need and you could help me at no cost to yourself. But suppose there is someone else in even *greater* need whom you could instead help at no cost to yourself. Suppose you cannot help us both. You have a pro tanto duty in one respect (need) and to an extent (based on the degree of need) to help me, but you also have a pro tanto duty in the same respect but to a greater extent to help the other person. Thus, your all-things-considered duty in this situation is to help the other person.

The term 'duty' is more often used and heard in the pro tanto sense than in the all-things-considered sense. The example in the previous

paragraph concerned a conflict of *pro tanto* duties. But we could have expressed the example in terms of a conflict of *duties*. Saying that you have a duty to do something in a situation is not to proclaim an all-things-considered moral verdict about what to do in that situation, since you might also have an opposed duty in this situation to do something else. I am not denying, however, that we can also use the word 'duty' to mean an 'all-things-considered moral verdict about what to do'. Indeed, precisely because the word 'duty' is ambiguous between 'pro tanto duty' and 'all-things-considered duty', Ross did the world a favour by distinguishing between these meanings and offering terminology to mark the distinction.

Just as 'duty' has a pro tanto sense and an all-things-considered sense, so does the term 'ought'. However, whereas 'duty' is normally meant and understood to be referring to pro tanto considerations, 'ought' is normally heard as affirming an 'all-thing-considered' judgement. Return to the example where I am in need and you could help me at no cost to yourself, or you could instead help someone else who is in even greater need than I am and your helping that person would impose no cost on you. If someone asserted that you ought to help me, you might reject that assertion on the grounds that there is another person in greater need whom you could also help at no cost to yourself. If others' needs are the determining factor, then it seems natural to say that really what you ought to do is help the other person. I submit that you would *not* be inclined to say that you are subject to conflicting oughts. Rather, you would be inclined to say instead that whatever in the end you should do is what you ought to do, with the outweighed consideration's failing to be an ought at all.

The dominant moral theories in the 1950s and 1960s in Anglophone countries were utilitarianism and Kantianism. Kant's Categorical Imperative tells one what one ought morally to do—act on maxims that one can will to be universal laws; or treat others always as ends in themselves, not merely as means. The Categorical Imperative is not telling one merely what considerations to weigh up when making moral decisions. The simplest form of utilitarianism is also framed as an imperative—do whatever maximises utility, impartially calculated. Both Kant's Categorical Imperative and this simple form of utilitarianism were often framed in terms of 'ought' judgements. These were all-things-considered 'ought' judgements, not merely pro tanto ones.

Monistic moral theories offer a single imperative based on a single consideration, even if that single consideration is complex. The first formulation of Kant's Categorical Imperative, for example, identifies just one thing as determinative, namely whether the agent (arguably Kant meant *every* agent) could will her maxim as a universal law. And the simplest form of utilitarianism is an act-utilitarian theory holding that what one ought to do depends entirely on what maximises aggregate utility, where benefits and harms to everyone, including future people,

are included. Kantianism and this simple utilitarianism are alike in pointing to a single consideration, albeit not the same one. Each of these two theories does not need the pro tanto/all-things-considered distinction because each of these theories denies that there are different kinds of moral considerations to be weighed against one another.

I do not mean to deny that utilitarianism and Kantianism cannot make use of the pro tanto/all-things-considered distinction. For example, even the simplest utilitarianism might allow that each possible benefit generates a pro tanto reason to promote it, and these reasons are to be weighed together and against other such reasons in order to ascertain an all-things-considered moral verdict on what to do. Kantianism might have use of the distinction in other ways. My point was not that the distinction cannot be used by utilitarianism and Kantianism. My point was that act-utilitarianism and Kantianism do not need to use the distinction.

Having commented on Kantian and utilitarian theories, I turn to virtue ethics, by which I mean ethical theories that evaluate action as right or wrong by reference to what a virtuous person would characteristically choose. Does virtue ethics need the pro tanto/all-things-considered distinction?

The answer seems to me to depend on whether the virtues are conceived of as possibly conflicting with one another. One conception is that the virtues cannot conflict, because, for example, a person could not really grasp what kindness requires unless she also understood that kindness cannot require dishonesty. If the virtues cannot conflict with one another, then virtue ethics has no need of the concept of a pro tanto moral reason. In contrast, if the virtues can conflict with one another, then someone might find herself in a situation where, for example, kindness pulls in one direction but honesty pulls in the opposite direction. The pro tanto/all-things-considered distinction helps articulate such conflicts: There is a pro tanto reason to be kind and a pro tanto moral reason to be honest, and what the agent should do, all-things-considered, depends on what a virtuous person in this situation would characteristically take to be the more important moral reason when they conflict in the situation at hand.

During the 1940s, 1950s, and 1960s, most ethical theorists in Anglophone countries directed their attention to utilitarianism and Kantianism, on the assumption that both virtue ethics and moral pluralism are unsatisfactory moral theories. Virtue ethics was presumed to be either implausibly committed to holding that virtues cannot conflict or to be in effect a form of moral pluralism. The prevailing objection to moral pluralism was that it, in D. D. Raphael's words,

> does not meet the needs of a philosophical theory, which should try to show connections and tie things up in a coherent system. To

> look for unity where none exists would, of course, be foolish; if the diversity of moral rules were intractable, it would be pointless to go on searching for some way of tying them up together. But the moral rules of ordinary life are not obviously all different from each other.
> (Raphael 1994, 55)

A related objection to moral pluralism is that it is often less determinate in its practical applications than utilitarianism and Kantianism.

Nevertheless, the prevailing methodology in normative ethical theory since about 1970 seems to me to have elevated moral pluralism to the title of 'theory to beat'. This methodology holds that we should take our considered moral convictions at all levels of generality and try to get them into 'reflective equilibrium' with one another, and with everything else we believe. Thus, we expect our general moral principles to be compatible with our non-moral beliefs and to cohere with the specific moral convictions we have after due consideration.

The methodology of seeking reflective equilibrium is normally traced to John Rawls (1951; 1971, 19–21, 46–51; 1974–1975, sect. 2; 1980, 534). Rawls assumed that the pressure to find general principles is so strong that we should endorse whatever is the most attractive set of general principles we can find and we should discard moral convictions that do not accord with that set. I agree that, if two sets of general moral principles are equally good at supporting our more specific moral convictions but one of the sets of general moral principles provides this support on the basis of fewer principles, this more parsimonious set of principles is better. What makes it better is that it explains equally much on the basis of less. But we should not take the method of reflective equilibrium to be pre-committed to endorsing whatever *single* fundamental moral principle supports the highest percentage of our other moral convictions. For all we know prior to thorough investigation, there is some set of plural fundamental principles that (a) is consistent with our non-moral beliefs, (b) seems plausible, and (c) coheres with all our more specific moral convictions better than any single fundamental principle on its own does.

Indeed, despite the pressure exerted on us by having parsimony and connectedness as desiderata in our moral theorising, we might well think that the method of reflective equilibrium takes us to moral pluralism. Moral pluralism does an unbeatable job of agreeing with our various moral convictions. Unless some single fundamental moral principle can do as good a job of agreeing with our various moral convictions as moral pluralism does, then moral pluralism is the moral theory best justified to us (Hooker 1996; 2000a, ch. 1).

I return to the point that moral pluralism is best expressed using the pro tanto/all-things-considered distinction. The term 'normative reasons' is most naturally understood to be referring to normative favourers or disfavourers, i.e., pro tanto normative considerations rather than all-

things-considered verdicts about what to do. The emphasis on reasons makes sense if we need to acknowledge different, possibly conflicting normative pressures, of possibly different strengths, before we decide what to do. For many of us, everyday practical decision making is sprinkled with episodes of weighing up the strengths of conflicting pressures, leading up to a conclusion about what, all-things-considered, we should do. Thus, thinking in terms of reasons comes very naturally.

Rules

Trying to think only in terms of reasons and bypass rules entirely, however, would be disempowering. A great deal of thinking involves recognising a difference between things, which often forms the basis of reacting to different things differently. Glock writes, 'Judgement precisely involves classifying an object *as* being of a certain kind. And this idea can in turn be spelled out by saying that judgement requires the deliberate choice between different options in a sorting or discrimination task' (2010, 19). Without wading into the discussion of Glock's arguments about animal thought (which are taken up by other contributors to this volume), I note that, at least in the case of beings with language, to classify or categorise something as being an instance of a kind is a rule-governed activity.

According to Glock, 'The linguistic meaning of expressions depends on general rules. These rules provide standards for the correct use of expressions' (2015, 842). The rules providing standards for the correct use of expressions are *normative* rules shared by the linguistic community. With respect to the rules determining linguistic meaning, I have no reason to dissent from Glock's drawing on H. L. A. Hart (1961):

> [I]n a group G a behavioural regularity R is a *shared rule* if and only if
>
> 1 it is rare for members of G to deviate from R
> 2 if members of G deviate from R, they are subject to sanctions, including the verbal sanction of being criticised
> 3 these sanctions are generally accepted by members of G. (Glock 2019b, 313)

Much of what Glock takes from Hart can be applied to moral rules as well as to rules about linguistic meaning. When we think of possible moral rules, we think of rules to be shared in the sense that

1 People would comply with the rules and at least sometimes use the rules to guide their behaviour.
2 Non-compliance would be met with sanctions, such as blame, indignation, resentment, the withdrawal of good will and cooperation, etc.

3 Both the rules and their connection to sanctions would be accepted as justified.
4 People would think of the rules as generating reasons for action and grounding interpersonal justification.

One difference between linguistic rules and moral rules is that linguistic correctness is determined by linguistic rules that are *already* shared by those with linguistic competence. Already shared rules are hardly static. People can make proposals of linguistic reforms, coin new terms, and innovate in other ways. Nevertheless, linguistic correctness in general is determined by already shared rules.

When we turn from linguistic rules to moral rules, the fact that a rule is already shared seems to me less authoritative. Let us distinguish between a set of rules that are *already* shared and a set of rules that might not be already shared but *ideally* would be shared. There could be a society where the two sets of rules are the same—people there already accept the ideal rules. But the importance of the distinction between established rules and ideal rules comes out when we focus on a society in which the established rules are not ideal. For example, consider a society in which the *actually shared* rules insist that one should unwaveringly identify either as male or as female and then routinely behave very differently depending on which one is. But *ideal* rules would neither demand that one identify as being of either one of two genders nor require different behaviour depending on gender. Because of this point about ideal rules, refusing to identify as being of either one of two genders is morally permissible. Admittedly, there might be powerful self-interested reasons not to offend against the rules accepted by the people around you. But if the rules they accept are unnecessarily restrictive, or invidiously discriminating, or destructively lax, the bare fact that those rules are already accepted does not entail that they dictate moral correctness.

Rules articulating requirements generate reasons for action. For example, there is a moral rule requiring people not to steal, and this moral rule is one source of reasons not to steal. Requirements on action also lead to reasons to react to behaviour with (e.g.) feelings of indignation, resentment, or guilt and withdrawal of goodwill, ostracism, or more (see Scanlon 2013, 105–109; 2018, 120–121).

For rules, including rules making requirements, to be justified, there must be undefeated reasons for having these rules. Reasons for having these rules are more basic than the rules themselves: The reason for having the rules is not that there were already reasons for action that were independent of and prior to the rules. Rather, the reason for having the rules is that people's accepting the rules would have better consequences than people's not accepting them would—where 'better consequences' is not merely a matter of people's more often acting on reasons they already had. To justify these assertions about which reasons are basic, I would

have to digress for longer than tolerable. (But see Copp 2010; 2020; Parfit, 2017, 432.)

Moral rules not only impose requirements but also grant permissions and powers. For example, there is a moral rule permitting people to devote their own time, energy, attention, and other resources to achieving their own goals even when these resources could instead be used to help other people achieve their similar or even somewhat more important goals. Admittedly, this permission is not unlimited. For example, one might need to sacrifice one's goal of getting home early in order to rescue an accident victim whom one comes across on the way home.

The moral rule *permitting* people to devote their own time, energy, attention, and other resources to achieving their own goals does *not* itself give people *reasons* to do what the permission allows. To be sure, people *do* have reasons to devote their own time, energy, attention, and other resources to achieving their own goals. But these reasons do not come from the moral permission they have to do so. The point generalises: Permissions to do things do not *on their own* generate reasons to do what the permissions allow.

Permissions are liberties. If holding on to a particular liberty is important and if occasionally exercising the liberty is necessary to hold on to it, this might give you a reason to exercise the liberty, though not necessarily an undefeated reason. What gives you the reason in such a case is not the liberty *on its own* but rather the combination of the value of preserving the liberty and the necessity to exercise the liberty occasionally in order to preserve the liberty. Nevertheless, a person's permissions and liberties *do* entail *reasons* for action *for other people*. If a person is at liberty to do a kind of act, other people have reason not to force that person not to do that kind of act.

There are also moral rules specifying normative powers (Owens 2012; Hohfeld 2019; Chang 2020). Suppose that up to now you have no right against me that I meet you tomorrow for lunch. But if I now promise you I will meet you tomorrow for lunch, I have thereby created a moral obligation on me and a moral right in you against me that I meet you tomorrow for lunch. Your new right comes with your new moral power to waive your right and cancel my obligation.

The example of promising as a social practice constituted by rules brings out one of the ways in which rules are ineliminable from morality as we know it. (On the kind of rules that constitute a social practice, see Rawls 1955; Glock 2019b, 307.) Many moral actions make sense only when construed as complying with a social practice constituted by rules. To explain what a promise is, we have to refer to rules about agents' voluntarily creating obligations for themselves and rights for others.

Rules constitutive of social practices are not the only ones to pervade moral thinking. Suppose you are trying to decide what to do in a situation in which many innocent people would benefit or be harmed by

your decision. Let us add that some of the alternative acts you could choose would involve varying degrees and kinds of dishonesty. We already know enough about the situation to see that various pro tanto moral reasons come into play. One is a pro tanto moral reason to benefit innocent others, with the more benefit to others and the more others, the better. Another is the pro tanto moral reason not to harm innocent others. And another is the pro tanto moral reason to avoid being dishonest.

While each of these moral reasons is one that *you* have *in this particular situation*, they are instantiations of general phenomena: *All* agents have pro tanto moral reasons to benefit innocent others, not to harm innocent others, and not to be dishonest *in all situations in which acts of these kinds are available*. And such *general* moral reasons are often expressed as rules in the imperative voice—such as the rules 'do good for others', 'don't harm others', and 'do what is honest'. (On the 'inherent generality' of rules, see Glock 2015, 843.)

A point regularly made is that common knowledge that certain rules are widely accepted plays an immensely important role in coordinating people's behaviour. A standard example is that each society needs there to be common knowledge about which side of the road to drive on there. Not every rule solves a coordination problem (Glock 2019b, 312), but many do, and many coordination problems would be very difficult to solve without shared rules.

Perhaps even more importantly, common knowledge that certain moral rules are widely accepted also plays an immensely important role in assuring people about other people's behaviour. Admittedly, we know the law forbids (e.g.) stealing, and law enforcement attempts to deter stealing. Nevertheless, we will be even more confident that others will not steal from us if we believe that, in addition to others' awareness of the law and the punishment imposed for stealing, others accept a moral rule against stealing. What I have written here about stealing also applies to physical aggression and promise breaking. Moral and legal rules had better 'protect persons, property, and promises' (Hart 1961, 193).

Could human societies realistically make do with thinking in terms of reasons and eschew all references to moral rules? I suspect that reducing all moral rules to propositions about reasons could not be done without losing valuable information. However, even if such reduction is possible, I expect that many of the resulting propositions about reasons would need to be long and complicated, thus sacrificing the simplicity that many rules have. Sacrificing that simplicity would have high costs in terms of cognitive overload and the corrosion of assurance (see Hooker 2000b, 2007).[1]

Note

1 In September 2021, this paper was presented at a conference in Hanjo Glock's honour. I thank Christoph Pfisterer, Nicole Rathgeb, and Eva Schmidt for

very helpful comments on a draft given to me prior to the conference. I also thank Maria Alvarez, Gerhard Ernst, Hanjo Glock, John Hyman, Jöerg Löschke, and Christoph Pfisterer for comments given to me at the conference. Each of these people made comments that significantly improved the paper.

References

Chang, R. (2020). Do We Have Normative Powers? *Aristotelian Society* 94: 275–300.

Copp, D. (2010). The Wrong Answer to an Inappropriate Question? In Black, S. and Tiffany, E. (eds.), *Reasons to Be Moral Revisited*. *Canadian Journal of Philosophy* 33: 97–103.

Copp, D. (2020). The Rule Worship and Idealization Objections. *Oxford Studies in Normative Ethics* 10: 131–155.

Glock, H.-J. (2010). Can Animals Judge? *Dialectica* 64: 11–33.

Glock, H.-J. (2015). Meaning and Rule Following. In *International Encyclopedia of the Social & Behavioral Sciences*. 2nd edition. Vol. 14. Oxford: Elsevier, pp. 841–849.

Glock, H.-J. (2019a). Agency, Intelligence and Reason in Animals. *Philosophy* 94: 645–691.

Glock, H.-J. (2019b). Normativity of Meaning Revisited. In Roughley, N. and Bayertz, K. (eds.), *The Normative Animal?* Oxford: Oxford University Press, pp. 295–318.

Hart, H.L.A. (1961). *The Concept of Law*. Oxford: Clarendon Press.

Hohfeld, W. (2019). *Fundamental Legal Conceptions*. Edited by W.W. Cook. New Haven, CT: Yale University Press.

Hooker, B. (1987). Williams's Argument against External Reasons. *Analysis* 47: 42–44.

Hooker, B. (1996). Ross-Style Pluralism versus Rule-Consequentialism. *Mind* 105: 531–552.

Hooker, B. (2000a). *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford: Clarendon Press.

Hooker, B. (2000b). Moral Particularism—Wrong and Bad. In Hooker, B. and Little, M. (eds.), *Moral Particularism*. Oxford: Clarendon Press, pp. 1–22.

Hooker, B. (2007). Moral Particularism and the Real World. In Lance, M., Potrc, M. and Strahovnik, V. (eds.), *Challenging Moral Particularism*. London: Routledge, pp. 12–30.

Kiesewetter, B. (2017). *The Normativity of Rationality*. Oxford: Oxford University Press.

Lord, E. (2018). *The Importance of Being Rational*. Oxford: Oxford University Press.

Nagel, T. (1970). *The Possibility of Altruism*. Oxford: Clarendon Press.

Owens, D. (2012). *Shaping the Normative Landscape*. Oxford: Oxford University Press.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Clarendon Press.

Parfit, D. (2017). *On What Matters*. Vol. 3. Oxford: Oxford University Press.

Raphael, D. (1994). *Moral Philosophy*. Oxford: Oxford University Press.

Rawls, J. (1951). Outline for a Decision Procedure in Ethics. *Philosophical Review* 60: 177–197.

Rawls, J. (1955). Two Concepts of Rules. *Philosophical Review* 64: 3–32.
Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Rawls, J. (1974–1975). The Independence of Moral Theory. *Proceedings and Addresses of the American Philosophical Association* 48: 5–22.
Rawls, J. (1980). Kantian Constructivism in Moral Theory. *Journal of Philosophy* 77: 515–572.
Ross, W.D. (1930). *The Right and the Good*. Oxford: Oxford University Press.
Scanlon, T. (2013). Giving Desert Its Due. *Philosophical Explorations* 16: 101–119.
Scanlon, T. (2018). *Why Does Inequality Matter?* Oxford: Oxford University Press.
Sidgwick, H. (1907). *Methods of Ethics*. London: Macmillan.
Williams, B. (1981). Internal and External Reasons. In Williams, B. (ed.), *Moral Luck*. Cambridge: Cambridge University Press, pp. 101–113.

16 No reason to be afraid! On the (ir)rationality of emotions

Gerhard Ernst

Among the many things I have learned from Hanjo Glock over the years, one is to keep an open mind: If you think Quine, Davidson, and Wittgenstein are great philosophers, it does not mean that you have to agree with them. Better take a look at their blind-spots! If you think, after years of practicing it, that you know pretty well what analytic philosophy is, make sure your opinion is not the result of a one-sided diet of examples. Take a look at what analytic philosophy has *also* been! Hanjo Glock always goes the extra mile to bring into focus overlooked sides of philosophical positions and problems.[1] For this reason, it is quite hard to find an aspect of a topic that he has worked on whose importance he may have underestimated. Hanjo Glock is very interested in the concepts of reasons and rationality but, at least to my knowledge, he has not published on reasons for or the rationality of emotions. I present this chapter to him in the hope that he will find in it something valuable concerning the general topic of reasons and rationality.

My wider project is to contribute to answering the question of how to understand reasons and rationality in general. But in the present chapter, I am going to focus on the following sub-questions: Under which circumstances are emotions rational or irrational? And what are reasons for emotions? In the main body of the chapter, I will discuss seven mini cases in order to shed light on these questions. I will start, though, with a few words on what I mean by 'emotion' and a brief explanation of why answering the questions of this chapter is of theoretical as well as of practical importance.

Quite a lot has been written on what emotions are,[2] but I will mainly try to steer clear of that literature in what follows and avoid being drawn into debates on the nature of emotions. I want to be committed only to what is firmly implied by our ordinary use of the word 'emotion' (and to what is, therefore, (pretty) uncontroversial in these debates), and to present an account of the (ir)rationality of emotions which should be compatible with all or at least most accounts of emotions. The philosophical discussions of the concept of emotion go well beyond our ordinary use of the word 'emotion': They often aim at an explication, not just an analysis of the concept.[3] And when it comes to explication, everything depends

DOI: 10.4324/9781003202929-21

on the—practical and theoretical—purposes we have. Philosophers giving an account of emotions today often try to accommodate empirical findings from other disciplines such as psychology, neuroscience, and evolutionary biology. I do not want to engage in this kind of project in the present chapter, since I am more interested in the general concept of rationality than in the concept of emotion, and I do not want to provide a 'theory of emotions'. Therefore, I am mainly going to focus on the rationality and irrationality of clear everyday examples of emotions, the emotion of fear in particular. The account of the (ir)rationality of emotions I want to present is meant to apply to this and similar 'folk emotions'.

This project is of theoretical interest, since it provides part of the answer to the question about the nature of reasons and rationality in general. There are lots of papers on the rationality and irrationality of beliefs and intentions. Less has been published on the rationality of emotions. And it is easy to see why: At least at first glance, there is an opposition between the sphere of emotions on the one hand and rationality on the other. This opposition is present in philosophical discussions of emotions since antiquity and is often alluded to in popular culture. The opposition between 'the head and the heart' in particular, i.e. the opposition between reason and love (and other emotions) is the psychological backdrop of countless books and films. The idea is usually that you *either* follow your heart (i.e. your emotions) *or* your head (i.e. reason). Following both is impossible, since emotions seem to be per se irrational (or at least a-rational), whereas reason is 'un-emotional'. There is no room for rational emotions here. Hence, the popular answer to our question—under which circumstances are emotions rational—is either 'None! Emotions are always irrational' or cannot be given (if emotions are, by their nature, neither rational nor irrational). You are either an emotional or a rational person. A 'rational emotion' looks like an oxymoron.

But this impression is, of course, misleading. The popular opposition between emotion and rationality is not due to the allegedly irrational or a-rational nature of emotions but is rather a consequence of the fact that rational emotions are not the stuff from which good stories are made. If rationality and emotion point in the same direction, it easily becomes boring.

There are more serious reasons for believing that emotions are, by their nature, neither rational nor irrational. We have emotions, one might think, or we do not have them, just like headaches. And just like there is no such thing as rational or irrational headaches, there seems to be no such thing as rational or irrational emotions. It is useful to distinguish two thoughts here:

One thought is that emotions are as independent of our will as headaches are. And that is the reason why they are neither rational nor irrational. This is not convincing for two reasons: The first reason is that it is by no means clear that emotions are *totally* independent of our will. If

someone is in a panic, for example, it is not pointless to say something like 'Calm down!' (whereas it would be pointless to urge someone to stop having a headache). We do have some kind of control over emotions. Perhaps we can, within limits, decide to calm down, whereas we lack any direct control in the case of pain. The second reason is that voluntariness is no precondition of rationality. Beliefs can clearly be classified as rational or irrational, and they are as hard to control voluntarily as emotions.

A second thought is that emotions, like heartache, and bodily feelings, like a headache, share something important: They both have a phenomenal aspect: They (typically sometimes)[4] feel a certain way. Is this not a reason for classifying them neither as rational nor as irrational? I think this is an open question. Emotions typically sometimes do have a feeling quality, so much so that they can even be accompanied by bodily sensations. Too much heartache cannot only give you a headache but can really feel like a pain in the heart. But whether this is a reason to exclude emotions from the realm of rationality is yet to be seen. And if they should not be excluded, they are part of what, in turn, shapes our concepts of reasons and rationality.

The question about the rationality of emotions is not only of theoretical but also of practical importance. I do not want to assume here that it is *always* good to be rational and that, therefore, we should always have rational feelings, and *always* bad to be irrational, and that, therefore, we should always avoid irrational feelings. And I also do not want to assume that we can choose our emotions just like we can choose our actions. But I do think that it is very often good to have rational emotions and very often bad to have irrational ones, and also that we can do a lot, maybe directly and certainly indirectly, to manage our emotions. It is of practical importance to know under which circumstances our emotions are rational and under which irrational, since the rationality or irrationality of our emotions is relevant to how good our lives are. I will later on have a brief look at how emotions can be good or bad for our lives.

Another practical question in this connection—one I only want to mention in order to lay it aside—is under which circumstances one is allowed or maybe even required to point out to someone that her or his emotions are rational or irrational. It is one question whether the emotions of another person are, in fact, irrational or not. It is quite another question whether you are entitled or maybe even required to say so. Prima facie there is every reason to believe that the emotions of another person are a personal matter, and that they are none of our business. Our emotions are intimately connected to our self-image and also to our feeling of self-worth. The evaluation of the emotions of someone, therefore, tends to be part and parcel of the evaluation of the person as a whole, more so than, for example, the evaluation of her or his beliefs or even actions. Also, our emotions (unlike many of our actions) often

do not have immediate consequences for other people. And when they do not, one of the main reasons for justified external critique does not apply. Finally, one should keep in mind that the allegation of the irrationality of emotions is often a means for manipulating people. Therefore, the step from judging an emotion of another person as rational or irrational to saying so has to be taken with great ethical sensitivity. And I will remain silent on this topic in the present chapter.

I want to discuss the question under which circumstances emotions are rational or irrational on the basis of a series of little cases. They will, I hope, bring out clearly the main aspects of the (ir)rationality of emotions. Here is the first one.

> Case 1
> John believes there is a logic examination tomorrow he has to attend. Therefore, today he is very afraid of tomorrow's examination. His friend Abbie tries to calm him down: 'You don't need to be afraid, John, since the examination has been cancelled due to the corona crisis.'

We would, just like Abbie, expect John to calm down after learning that there is no examination tomorrow. His fear should vanish. If John were still afraid, his fear would be an irrational emotion because John would be afraid of something (an examination) even though he knows that it will not take place. Not only would John's fear be objectively[5] without reason, if Abbie is right; his fear would also be irrational, since he knows that it is without reason. The irrationality of his emotion would manifest itself in the fact that neither Abbie nor John himself would understand why he was still afraid. Rational fear is directed at certain (real or envisaged) things.[6] If you do not believe that the things you fear will happen, your fear is irrational.

Concerning the rationality of John's fear, it does not matter whether Abbie is, in fact, right about the examination: It only matters whether he believes her. If, in fact, the examination takes place after all, objectively John would have a reason to be afraid. But, as long as he does not believe that this reason obtains, his fear is irrational. Things would be different if John did not believe what Abbie tells him, that is, if he continued to believe that there is an examination tomorrow. Maybe this belief would then be irrational (if John sees no reason to distrust Abbie), but his fear would not be irrational (given his belief). Rationality supervenes on the mind—in the case of emotions as well as elsewhere.[7]

So, this is a first case of irrational emotions. Here is another:

> Case 2
> Again, John believes that there is a logic examination tomorrow, and therefore he is very afraid today. His friend Abbie tries to calm

him down by saying: 'Look, you don't need to be afraid, since you are well-prepared for this examination.'

Again, we would, like Abbie, expect John to calm down, that is, we would expect his fear to vanish or, at least, be diminished. If John replied: 'I am still *very* afraid', his fear would be irrational. Why? What John is afraid of, the upcoming logic examination, actually takes place, at least for all he knows. Again, it is irrelevant for the rationality or irrationality of his fear whether the examination actually takes place or not. What is relevant is whether John believes it to take place (and let us assume that it does in this example). This time, Abbie does not try to calm John down by telling him that what he fears will not happen. Rather she tries to show that it is no threat (or not as big a threat as John thinks it is). There is no reason to fear a logic examination if you are, like John, well-prepared. Abbie does concede implicitly that logic examinations are a threat for the ill-prepared. But studying logic, i.e. preparing for the examination, is an adequate means for dealing with this threat. A lion is a dangerous animal. But if it is behind bars, the threat is averted. According to Abbie, John has averted the potential threat of the logic examination by preparing appropriately. And just like it is irrational to fear something which will not happen, it is irrational to fear something which is no threat (any more). Rational fear is directed at (really or supposedly) dangerous or harmful things.[8]

Just like it is irrelevant for John's rationality whether there is, in fact, an examination—it only matters whether he believes this to be the case—it is irrelevant whether the thing he fears is, in fact, dangerous or harmful—it only matters whether he believes this to be the case. In case 2, I assume that John and Abbie agree that a logic examination is a threat only for the ill-prepared, and that you have nothing to fear if you are well-prepared. If the examination were, in fact, difficult to pass even for the well-prepared, John's fear would be justified objectively. Still, he would be irrational in fearing the examination as long as he believes that there is no risk of failing the examination for him (because he is well-prepared). (And again, John's fear would not be irrational if he believed that the examination is a threat even for the well-prepared, although this belief, in turn, might be irrational. But that is another matter.)

We would, therefore, expect John to justify his fear perhaps like this: 'Yes, Abbie, you are right, I am well-prepared. But I am still afraid that I might have a panic attack during the examination, and then fail the examination even though I have studied hard.' We could understand John's fear because it is now directed at something which is, indeed, a real threat: a panic attack during the examination. The means to block the initial threat, i.e. preparing for the examination, seems to be unreliable. It is as if the lion is locked in a cage with rusty bars. If, on the other hand, John says: 'Yes, Abbie, you are right: I am well-prepared, and I

know that everything will go well. But I am still afraid', we cannot fully understand his fear, and neither can he. It would be an irrational fear.[9]

What if John is not well-prepared? If he believes that the examination takes place and also that it is difficult to pass for the ill-prepared, his fear would not be irrational. His fear would fit the reasons as perceived by him: A difficult examination is a reason to be afraid (if failing the examination has harmful consequences). Or is it? Is there ever a good reason for fear? I will return to this point later. But let us note here that we can easily say that there is *no* reason to be afraid if the purportedly dangerous state of affairs does not obtain or is not dangerous.

Consider a third case:

> Case 3
> Again, John is very afraid of tomorrow's logic examination. Abbie tries to calm him down: 'Come on, John, it is only a logic examination! You'll live!'

We can assume that Abbie would consider logic examinations exciting or upsetting, especially if you are ill-prepared. But that is it. Even if you are not prepared at all, it is still only an examination. And the consequences of failing an examination are (usually) not *very* harmful. Therefore, there is no big *danger*. It is quite understandable that you are a bit tense or nervous or even a little scared if you have to take an examination. But having fear or being very afraid is plainly irrational. If John agreed 'Yes, Abbie, you are right, it is just a silly examination. But still I am terribly afraid', he would be irrational. Rational fear is directed at (really or supposedly) dangerous things, and the extent of the fear fits the extent of the danger.[10]

Using the concept of a reason we could describe the example like this: John's fear is irrational because there is (even from John's point of view) no reason to be *very* afraid, even though there is reason to be a bit nervous. John's feelings are not in tune with the reasons as he sees them.

Usually, we do not fear things we believe not to happen, and we also do not fear things we do not consider to be dangerous. But it is quite common to be more afraid than one thinks is warranted. And it also happens that we are less afraid than we think we should be. In the first case, we speak of cowardice; in the second case of recklessness. I think, both are paradigm cases of irrationality even if one does not *act* accordingly: The relevant emotions themselves are irrational.

There is another, perhaps rare, case of irrational emotions. It looks like this:

> Case 4
> This time John believes that the examination is a real threat. But somehow he is amused by this. Abbie, who is afraid of the

> examination herself, says: 'A logic examination is a serious matter. What's there to laugh about?'

John's emotion, mirth, is irrational twice over. Not only does John fail to fear what he considers to be dangerous; he is also amused about what he does not consider to be funny. Of course, if John believed logic examinations to be risible, we could understand his amusement. It is only because he, just like Abbie, believes it to be a threat that his mirth is inexplicable, even for himself. Using the concept of a reason, we could say: In this case not only is there (in the eyes of John) reason to be afraid; there is also (in the eyes of John) no reason to be amused. But, again, one might ask: Is there ever reason to have fear? And is there not always reason for cheerfulness? We will return to this point.

Let us consider another, maybe even rarer, example of irrational emotions:

> Case 5
> Again John is afraid because of tomorrow's logic examination. This time Abbie is totally surprised: 'Why are you afraid now? You have been telling me over and over again how much you are looking forward to this examination!'

What happens here? Is it possible for John to be afraid of what he looks forward to?[11] It seems to be impossible to *feel* fear and pleasure concerning something at the very same time. But you can experience both feelings in quick succession. It might be that John now is afraid of what he was looking forward to just a moment ago and will be looking forward to again in a moment—just to return to panicking again briefly afterwards. This may happen even if John's beliefs do not change at all. It rarely happens concerning logic examinations. But it does happen in our feelings towards other people. There is something like 'love-hate', and this emotion might just be such a mix of ever-changing feelings: Now we are drawn to someone, now we are repelled. But it is difficult to give a clear analysis of this emotion: Are we changing our beliefs (without reason), so that we consider a person now loveable, now disgusting? Or do we keep our beliefs and just our emotions change (without reason)? Or does it help to describe our emotion dispositionally: Since love is not a feeling at a specific time but rather a disposition to feel, think and do different things on different occasions, we could perhaps say that it is possible to sometimes feel hatred concerning someone one still loves. Or is this just not love?

I am not sure whether we have here a special case of an irrational emotion, one in which the irrationality lies in the tension between different feelings at different times, where the feelings, normally, are manifestations of incompatible emotions. Maybe we should ask: Would a certain

feeling be understandable for the person herself, given the other feelings the person has? If not, the emotion would be irrational.

Before turning to two final cases, I want to consider the previous cases together and put them in a wider context. First of all, we have to distinguish two conceptions of rationality which are usually seen as alternatives but which, I think, are both important components of a complete picture of rationality.[12] According to the first conception, rationality is always a matter of consistency among or fit between different mental states.[13] My own interpretation of this conception looks like this: Beliefs and intentions (and certain emotions too) are not occurrences but are connected to certain dispositions. The manifestations of these dispositions are criteria for ascribing the beliefs and intentions (and certain emotions). If someone, for example, believes that his name is John, he will in suitable circumstance answer the question 'What is your name?' with 'John', he will write 'John' when signing something, he will say that his name has four letters when asked, he will step forward when money is given to everyone called 'John', etc. There is an inconsistency or misfit concerning a belief if most criteria of ascribing the belief are met but some criteria for not ascribing the belief are met too. If, for example, someone believes that it is best to get out of bed at 7 o'clock, the person will say so when asked, she will advise other people to do so, she will set the alarm clock for 7, etc. But if this person does not get up at 7 (even though she is awake), this casts doubt on her belief. And if the person thought: 'One should get up at 7, but it is just so cosy here; just another 5 minutes [. . .]', the person would plainly be irrational, in this case: weak-willed or self-deceived about what she really believes. Not getting up, that is, not acting appropriately, is inconsistent or does not fit in with the person's belief—or the belief is absent, but then the fact that the person honestly says otherwise etc. is a sign of self-deception. In both cases, the irrationality of the person lies in an inconsistency or misfit between criteria for ascribing a belief.

When we take a look at our cases against this backdrop, we see that there is always an inconsistency or misfit of this kind present: The fear does not fit the belief that there is no danger or the extent of the fear is ill-suited to the belief about the extent of the danger. The fact that someone is afraid under certain circumstances is a reason to attribute to her the belief that there is some danger present in these circumstances. If the person lacks this belief, we are confronted with an inconsistency of the kind I have described. John could say: 'I don't really believe that I am in danger. But I am still frightened' or 'I don't really believe that the situation is very frightening. But I am still in panic'. The locution 'not really' signals the tension. And it is this tension in which the irrationality consists.

The reason why my account of rationality is not committed to any particular account of emotions is that the only difference a change in

the analysis of emotion would make would pertain to the location of the tension: If you take occurrent feelings as central for emotions, the question is whether these feelings fit in with the relevant (empirical and evaluative) beliefs and intentions of the person. If you do not think of emotions as occurrences but rather as dispositions to experience certain feelings, examples like 'love-hate' become intelligible: There is a tension between feelings at different times, and therefore it is difficult to ascribe one emotion. If you take emotions to be intimately connected to or partly constituted by beliefs—e.g. fear as a certain feeling in combination with the belief that some danger is present—then the tension lies *within* the emotion, as it were, i.e. in the 'feeling-belief', or between the 'feeling-beliefs' and intentions of a person. And if you think that there is a particular connection between emotions and motivation, the tension in irrational emotions can be located within 'emotion-intentions', or between 'emotion-intentions' and (empirical and evaluative) beliefs of a person. Whichever way you look at it: The irrationality of emotions can be explained in terms of an inconsistency or tension within or between mental states.

These considerations also explain our talk of reasons for emotions: Certain emotions fit in with certain beliefs (and intentions) but not with others. And this is the basis for talking about reasons for emotions. John's fear fits in with his belief that an examination is imminent (given that John sees examinations as a threat). The fact (if it is a fact) that there is an examination is a reason for fear—if examinations really are a threat. There is not really a reason for fear if there is either not really an examination imminent or examinations are not really a threat. But the rationality of John's emotion only depends on his beliefs concerning the matter. The general point that reasons are facts can in this way be combined with the idea that rationality is concerned with the consistency of mental states.[14]

I think that there is a strong analogy here between reasons for emotions and reasons for beliefs. Concerning reasons for beliefs, in my opinion, the crucial question is also whether certain beliefs are consistent with certain other beliefs (and inconsistent with still others). In any case: Whether or not we have a reason for a certain emotion or belief does not depend on the question of how *good* that emotion or belief is. Our worry whether there is ever reason for fear or against cheerfulness feeds on the idea that fear is unpleasant whereas cheerfulness is pleasant. But this is simply not relevant when it comes to reasons for emotions. Or is it? I will presently return to this point.

But let me first take a brief look at a second account of rationality. According to this account, the key concept for understanding rationality is not the concept of consistency (in a wide sense) among or fit between mental states. Rationality is rather to be understood as correctly responding to reasons.[15] I do think that this is a *second* important

conception of rationality (cf. Ernst 2020). And it is also relevant for understanding the rationality of emotions if certain evaluative beliefs, for example, the belief that examinations are a threat and therefore in a certain way bad, are *part* of what an emotion is. The reason is that rationally forming evaluative beliefs on the basis of certain empirical facts is, on my account, best understood as a form of correctly responding to reasons. If, for example, you fail to form the belief that something bad happens (at least in one respect) even though you do correctly believe that pain is inflicted on someone, you fail to respond to a reason (i.e. a potential reason for action), and that is a failure of your rationality. According to this model, the case in which John is afraid because he incorrectly believes that examinations are a threat would be a case of not responding correctly to reasons (or rather incorrectly responding to non-reasons) and therefore a case of an irrational emotion. If, on the other hand, evaluative beliefs are not part of emotions, I would say: The evaluative belief itself is irrational (in the sense of being formed as an incorrect response to reasons) but the emotion is rational (since it fits the false evaluative belief).

Let me turn to two final cases, which are, at least at first glance, of a different nature than the cases discussed so far.

> Case 6
> Again, John is afraid of the logic examination. This time Abbie tries to cheer him up like this: 'Oh, John! Vivir en miedo es cómo vivir a medias—a life in fear is only half lived. What's the point of being afraid? Does it help you pass the examination tomorrow? On the contrary: It only prevents you from realizing your full potential if you are afraid. You will get nervous and make more mistakes than if you stay calm. Man up! Grab the bull by the horns! Show them what logic is!'

Probably, most of us have tried to calm people down and have been calmed down like this. In our present context, the question is: Is Abbie saying that John's fear is irrational? She is saying that John's fear is not *good* for John. But does this mean that there is a reason against it? At least at first sight this is a clear example of 'the wrong kind of reasons'.[16] That an emotion is not useful for me does not make it inappropriate or irrational. If I face a lion (outside of a cage), I have, it seems, every reason to be afraid, even if my fear does not help me one bit or even diminishes my chances of survival.

Of course, one can interpret Abbie in another way: Abbie does not want to say that John's fear is irrational. She only wants to motivate him to *do* whatever he can to get rid of it—precisely because this would be good for John. It is possible, after all, to manipulate emotions: John might direct his attention towards his good preparation or he might seek

distractions from all thoughts concerning the examination. Perhaps his fear would thereby be diminished. During the examination, he might think of his love for Abbie whenever fear attacks etc. This might lead John, in the end, to be irrational with respect to the emotions he has: He might cease to be afraid of what he considers to be dangerous. But then, we might, at least sometimes, have good reasons to evoke irrational emotions or to overcome rational emotions.

In case 6, Abbie has pointed out the instrumental disvalue of John's fear. But emotions often are pleasant or unpleasant, and therefore also have intrinsic value or disvalue. Therefore, a final example might look like this:

> Case 7
> Abbie fights John's fear by saying: 'Look John, life is short, and we therefore have to make the most of it and enjoy it as much as possible. Forget the examination tomorrow and enjoy the present!'

Again, we could say that Abbie is not criticising John's emotion as irrational. She does understand John's fear. But again, Abbie wants John to do what he can to get rid of his fear. And that is because fear is an unpleasant emotion. Therefore, you always have at least one reason to overcome your fear, and you always have a reason to get yourself into a happy mood. But these are reasons for actions, not reasons for emotions, one could say. And if one took this view of the situation, one could stick to the account given above: The rationality of emotions is only a matter of fit or consistency.

I think this 'wrong kind of reasons'-manoeuvre is justified. Nevertheless, I do believe that an important part of the picture is still missing. And we can see what is missing if we ask ourselves: What is, for example, (at least in many cases) the best means to get rid of fear? Answer: Get rid of the danger! What is the best means to effect cheerfulness? Get in a happy situation! I do not want to be overly didactic, but the easiest way to diffuse fear of an examination is to study up! And this does not seem to be just a weird coincidence but has to do with the (or one) use that emotions have in our lives.[17] Many emotions seem to motivate actions or at least to bolster our motivation. Fear supports a motivation for a fight, flight, or freeze reaction. And you fight harder, run faster, or keep quieter if you are afraid than if you only abstractly believe that some danger is present. If you are outraged, you change the circumstances. And you do it faster and more consequently than if you only abstractly think that something is wrong. If you are cheerful, you stay at the party. And you stay longer than someone who believes but does not feel that the party is fun. And so on.

It is easy to think of an evolutionary explanation for the emergence of emotions: Those of our ancestors who were afraid enough were more careful and lived longer; those who were more cheerful stayed longer at

parties and had more offspring etc. But be that as it may: Many emotions do, in general, support our motivation to act in a useful way. Not always, of course. Fear can paralyse when action would be called for; rage might blind us where temperance would be necessary. But in these cases, it is not the emotion per se but too much (or too little) of it that causes problems. Too much fear paralyses, too much rage makes us blind, but the right amount of both seems to be very useful to us. And sometimes even paralysing fear and blinding rage are useful: when a freeze reaction is the only chance of survival or when acting out with all your might is the only way of defence.

Based on these considerations, we can ask: What constitutes the *right* amount of an emotion? Which degree of fear is appropriate to which degree of danger? Which amount of indignation is appropriate to which amount of wrong? I think a plausible answer to these questions is: The intensity of an emotion is appropriate if and only if it supports the motivation to act in an appropriate way—and is in this way consistent with it. In order to be rational, emotions have to fit evaluative beliefs (or fitting evaluative beliefs have to be parts of emotions). The motivation to act also has to be consistent with evaluative beliefs in order to be rational. Therefore, it seems natural to say that emotions are appropriate exactly if they, in turn, fit the motivation to act (or, again, fitting motivations have to be part of the emotion). In a rational person, evaluative beliefs, emotions, and motivations to act form a consistent, holistic state of mind.[18] If an emotion gets in the way of an intention, the emotion does not fit the relevant evaluative belief—given the intention fits the evaluative belief—and has to be classified as inappropriate.[19]

On the basis of these considerations, let us have another look at case 6 where Abbie asks John what he gains out of his fear and where she points out that his fear only makes him nervous and diminishes his potential. Given what I have just said, we can interpret Abbie as trying to point out the irrationality of John's emotion after all: She tries to show him that his fear is inappropriate to the situation. The problem is not simply that John's fear is useless for John or even detrimental to his aims per se, but that the degree of fear does not fit the degree of danger—given the appropriate motivations of John. It would be appropriate to feel a bit nervous, maybe have pre-exam jitters and a little stage fright at the beginning of the examination, and these emotions would help John in his motivation to prepare for the examination and later to focus on the task at hand. But outright fear simply does not fit the real 'danger' and hampers the appropriate actions. Therefore, we can interpret Abbie as really accusing John of irrationality, just like in the prior cases. If John's beliefs are right, that is, if he is not deluded about the danger a logic examination poses, his emotion is inconsistent with his beliefs and his appropriate motivations and therefore irrational.[20] When it comes to logic examinations, there really is no reason to be afraid.[21]

Notes

1 I was always particularly impressed by his two books: Glock 2003 and Glock 2008.
2 Cf. the extensive bibliography and overview in Scarantino and de Sousa's (2021) SEP article.
3 Perhaps, one would rather talk about 'conceptual engineering' today. For present purposes, it is not necessary to distinguish between 'explication' in Carnap's sense and 'conceptual engineering'. For the latter cf. Cappelen 2018.
4 It is often useful to distinguish between feelings and emotions, especially since some emotion terms (like 'love') refer to dispositions rather than mere occurrences. Still, it is pretty uncontroversial that emotions have at least a phenomenal *aspect* (i.e. typically sometimes feel a certain way). It is worth noting that the German translation of 'emotion' and 'feeling' is the same: 'Gefühl' (even though one can use 'Emotion' in German too).
5 I use the word 'objectively' here to mark the 'adviser perspective', i.e. the perspective of someone who knows all the relevant facts and is therefore in the position to offer perfect advice to the subject. It is hotly debated how best to account for this perspective in terms of reasons. Cf., e.g., Kiesewetter 2017, ch. 8.
6 Since it does not affect my main argument, I do not want to be committed to a specific ontological category of the proper object of fear—therefore, my talk of 'things' is meant in a loose sense.
7 For a discussion of this thesis, cf. Wedgwood (2002).
8 We can fear the *risk* (i.e. danger) of harmful things happening (e.g. the risk of failing an examination), but also harmful things themselves (e.g. the foreseeable pain of an operation).
9 This kind of irrationality might be quite common. Therefore, we might not be *surprised*. And, of course, there might be some causal explanation of the fear (and the fear then is in this sense understandable). Perhaps, we can also *partly* understand the fear if there is (from the point of view of the subject) *some* reason for fear. Still, we cannot *fully* understand the fear since, from the point of view of the subject, there is not *sufficient* reason.
10 Again: If John believed that the consequences of failing the logic examinations are really very harmful, his great fear would not be irrational.
11 Of course, it is possible to fear certain *aspects* of something while looking forward to other aspects of it.
12 I present this view in Ernst 2020.
13 This conception is elaborated by John Broome. Cf. Broome 2013.
14 The bullet one has to bite, if one adopts this view, is that the connection between reasons and rationality is loosened. It can be rational to be afraid even though one does not have a reason to be afraid.
15 Cf. Parfit 2011, for a detailed elaboration cf. Kiesewetter 2017.
16 For a discussion cf. Gertken and Kiesewetter 2017.
17 I am not making the stronger claim that it is the sole or core function of emotions or their very nature to motivate (as has been claimed by others; cf. e.g. Scaratino's and de Sousa's SEP article for an overview)—but what I say is compatible with the stronger claim.
18 In my opinion, it is best to view beliefs, emotions, and intentions as *aspects* of a holistic mental state. An irrational mental state is an inconsistent or strained mental state.
19 We act in order to change our environment in a way we judge to be good. According to our present consideration, we thereby automatically try to act

in a way that makes us feel good about our environment. That may be the reason why hedonists think that all our actions aim at making us feel good, that is, at pleasure. In my opinion, this gets things (at least partly) backwards: The emotion helps the motivation more than the other way round.

20 Because of its holistic character, the present account of the (ir)rationality of emotions sits uneasily with Scarantino and de Sousa's distinction between cognitive and strategic rationality (cf. Scarantino and de Sousa 2021, sect. 10).

21 I want to thank Nicole Rathgeb, Eva Schmidt, and Sebastian Schmidt for their very helpful comments on a previous draft of this paper.

References

Broome, J. (2013). *Rationality through Reasoning*. Chichester: Wiley Blackwell.

Cappelen, H. (2018). *Fixing Language. An Essay on Conceptual Engineering*. Oxford: OUP.

Ernst, G. (2020). Two Kinds of Rationality. In Schmidt, S. and Ernst, G. (eds.), *The Ethics of Belief and Beyond. Understanding Mental Normativity*. New York: Routledge, pp. 176–190.

Gertken, J. and Kiesewetter, B. (2017). The Right and the Wrong Kind of Reasons. *Philosophy Compass* 12: e12412.

Glock, H.-J. (2003). *Quine and Davidson on Language, Thought and Reality*. Cambridge: CUP.

Glock, H.-J. (2008). *What Is Analytic Philosophy?* Cambridge: CUP.

Kiesewetter, B. (2017). *The Normativity of Rationality*. New York: OUP.

Parfit, D. (2011). *On What Matters*. New York: OUP.

Scarantino, A. and de Sousa, R. (2021). Emotion. In Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition). https://plato.stanford.edu/archives/sum2021/entries/emotion. Last accessed: 28 March 2022.

Wedgwood, Ralph. (2002). Internalism Explained. *Philosophy and Phenomenological Research* 65(2): 349–369.

Part V

Reflections and replies

17 Reflections and replies

Hans-Johann Glock

I feel profoundly honoured by the attention lavished on my work by such esteemed colleagues and dear friends, chuffed by some of the exceedingly generous remarks, and moved by the reminiscences. It amounts to being overrated; but then, being overrated is much underrated. Thank you all, but especially the editors for devising this project and for its magnificent execution! I am almost overwhelmed by all the excellent questions, observations, and arguments assembled here. They go to show that 'the philosopher' *is* 'a citizen of a community of ideas', Wittgenstein notwithstanding (Z, § 455). Given the nature of this volume, I might, nevertheless, be forgiven for concentrating on those ideas that concern my own work.[1]

Reply to Severin Schroeder

I am gratified by Severin's formidable backing for my stance on hinge propositions, not least because the issues are far from straightforward. There is little to disagree with his judicious yet unequivocal assessment. Moorean truisms can be known, are empirical, and, taken in isolation, do not defy sceptical doubt. Some earlier passages in *On Certainty* suggest otherwise, yet Schroeder documents convincingly that Wittgenstein fought his way through to a more 'enlightened' position. I do detect the occasional relapse, however. OC, §§ 550–551, dated 18.4.51, no longer makes possession of knowledge by a subject *S* dependent on an answer to the question '*Why* does *S* believe it?', that is, on being able to adduce evidence. Yet it remains wedded to insisting that there must be an answer to '*How* does *S* know it?' And this answer is supposed to follow 'generally recognised principles'.

As Schroeder points out, often 'How do you know?' is answered by specifying a cause (loosely conceived) or *source* of my knowing, as in 'I was told by. ...' However, as Wittgenstein had come to recognise by the mid-thirties, not every kind of knowledge is based on a way or method of knowing. What is more, in some cases we cannot even adduce a source like a sense-modality or testimony. A prime example is a species of knowledge that Wittgenstein only took notice of late—proprioception

DOI: 10.4324/9781003202929-23

aka kinaesthesis (Glock 1996a, 192). I know that I am currently sitting. But being able to answer the 'How do you know?'-question is not a prerequisite of counting as a competent user of the pertinent expressions. Nor is it required for counting as a minimally reliable source of knowledge. Consequently, there are no 'generally recognised principles' that a subject of proprioceptive knowledge must have followed.

Schroeder also illustrates a point that I, among others, have tended to underestimate: the difference between an explanation that lays down a rule for using an expression, and an explanation through examples (22–24, 29). Both kinds of explanation combine with our conformity in judgements to provide a scaffold for our epistemic language-games.

Schroeder ends by pointing out that the real foundation of human knowledge is a practice manifesting and guided by standards of rationality. This is a salutary corrective to the still popular view that Wittgenstein's approach is fundamentally irrationalist. It is connected to another point, one that unites Wittgenstein with American pragmatism: Doubt requires reasons. It is here that scepticism is vulnerable. While purporting to implement a hyper-rationalist agenda, it lapses into silliness by failing to apply standards of rationality to the often tacit assumptions of sceptical doubts.

Reply to Joachim Schulte

Joachim recognises that our respective perspectives on Wittgenstein's thought are partly incommensurable, in that there is no argument settling all of the differences. He characterises my perspective as follows: 'Glock tends to see Wittgenstein in the context of a school of "analytic" philosophers' (34). In one respect this is too weak: I have gone out of my way in championing an analytic approach to Wittgenstein (Glock 2004, 2017c). In another respect it is too strong: I have been equally adamant that analytic philosophy is a highly diverse intellectual current or tradition rather than a school (Glock 2008b, chs. 5, 8).

Schulte applauds my distinction of different types of historicism for helping to address a vague subject matter in a more systematic fashion. At the same time, he complains that by using 'historicism' for 'any position which promotes historical thinking in philosophy and warns against ignoring or distorting the past' (Glock 2005, 238), the label loses its bite. In particular, according to Schulte it fails to capture Wittgenstein's sympathies for the more substantial historicism of Spengler (35–36). But my explanation of 'historicism' has the advantage of singling out historicism as an—admittedly general and vague—*methodological approach in philosophy*, while leaving aside historicism as a *philosophy of history* which regards the latter as a sequence of events governed by systematic and inexorable laws (teleological or cyclical). As Schulte nicely demonstrates, although Wittgenstein invoked the culture/civilisation

distinction in some of his reflections, he was *critical* of Spengler's substantive historicism. What Wittgenstein mainly garnered from Spengler was *not* a deterministic metaphysics of history, but methodological ideas revolving around the notion of an *übersichtliche Darstellung*. That is a topic that Schulte (1990) has written about authoritatively; yet it is not the topic of my (2005) article. The latter's explicit concern is with the question: What role does Wittgenstein accord to historical considerations and historiographical accuracy in philosophy, or should accord, given his overall thought?

My answer: Although Wittgenstein himself was entirely immune to the charms of historical scholarship, his conception of philosophy and of language supports a minimalist historicism, according to which conceptual historiography is helpful though not essential to philosophising (Glock 2005, 262). By contrast, Schulte diagnoses a 'much more substantial' kind of historicism, one which distinguishes and evaluates different periods, especially within European culture. Schulte rightly notes that none of my stronger variants of historicism fits this position either (17). The reason for this discrepancy is that our characterisations concern different topics: mine the role of history in Wittgenstein's discussions of *philosophical method*, his the relevance of historical context on the applicability of labels like 'great architecture' according to passages dealing with culture and value.

I find much to applaud in Schulte's discussion of these passages. I could not find any textual backing, however, for his claim that Wittgenstein treated 'spirit'—*in the sense of an attitude finding expression*—as primarily a collective phenomenon (37–38). Furthermore, Schulte's non-textual considerations to this effect (39–40) seem to run together the question of whether different individuals can *share* the same attitude with the question of whether *the primary subjects of attitudes*, shared or not, are individuals or collectives. Finally, if philosophising is to lead us 'back to the rough ground', then it had better start out with the attitudes of individuals rather than of collectives, let alone collectives like 'civilisations' or 'cultures' that are notoriously hard to pin down.

Reply to Daniel Whiting

Daniel defends a monistic interpretation of the later Wittgenstein's conception of nonsense according to which there is only privative nonsense, that is, nonsense resulting from our not having assigned a meaning to expressions in a certain context. By contrast to what I have called 'unruly Wittgensteinians', his case does not rest on denying that for Wittgenstein meaning is 'a matter of rules of use'. This is an interesting combination. In developing it, Whiting brings to light a host of important insights and challenges to pluralists like myself who ascribe to Wittgenstein the idea of combinatorial nonsense. Yet it is also an unstable combination.

Whiting usefully distinguishes 'top-down' and 'bottom-up' arguments for combinatorial nonsense. According to the former, combinatorial nonsense is implied by Wittgenstein's conception of meaning. Whiting acknowledges that the pertinent sense of 'nonsense' is that of linguistic unintelligibility rather than of pointlessness. Yet it is central to his case that while Wittgenstein 'does not identify the two, he takes there to be a close connection between intelligibility and pointfulness' (49). According to Whiting, it is a necessary condition for an expression being used meaningfully on a particular occasion that it 'provides or speaks to the relevant interests or needs' (53). One of his arguments in support of this interpretation fails. That the rules according to which we use expressions reflect our interests (52–53) does not entail that these expressions are used according to these rules only if their employment reflects these interests. Our social rules reflect (very imperfectly) our interests, but not every specific application of social rules does. Wittgenstein explicitly recognised the difference when he distinguished between what rules demand on the one hand, and the framework conditions which give point to adopting a rule on the other (Glock 1996a, 135–139).

Whiting's second argument carries more weight. In several places, Wittgenstein maintains that a school-grammatically well-formed sentence lacks sense since its use in a certain context fulfils no communicative purpose. This is a mistake on Wittgenstein's part—as Schroeder (this volume) and I (1996b) have argued. But that is beside Whiting's exegetical point. What is more relevant is that in other passages Wittgenstein keeps apart a sign-*type* being *meaningful* and the utterance of a *token* serving a *communicative purpose*. Even more pertinent to the top-down strategy: This equation is ruled out by Wittgenstein's idea that the meaning of a sign-type is determined by rules for its intelligible use. For *there are no* 'grammatical' aka semantic rules for the *effective* use of expressions. Potential rules for efficient communication, such as Gricean principles of conversation, span different languages and are not constitutive of the meanings of specific expressions belonging to particular languages. They are not part of acceptable explanations of meaning; and while their mastery is an essential component of human linguistic faculties, abiding by them is not a criterion of having understood specific expressions of a particular natural language.

Even if Whiting were entirely right on cases like 'I am here', this would not vindicate monism, since it leaves open the possibility that other cases of lack of sense are combinatorial. The real issue is whether the idea of an explanation of meaning implies the existence of combinatorial nonsense. In my discussion of

1 Caesar is a prime number

I had maintained that explanations like

2 'Julius Caesar' is the name of a Roman general

3 'is a prime number' is a predicate that applies to any number that can be divided only by 1 and by itself

'imply that the referent of "Julius Caesar" is not within the range of meaningful application of "is a prime number"' (Glock 2015, 124). Whiting retorts that they instead 'imply that the referent of Julius Caesar is not within the range of application of "is a prime number", that is, within its extension. So, what those rules imply is that (1) is false, rather than nonsense' (57). But such explanations specify *conditions* under which an expression refers or applies to something; they definitely neither state nor imply that, as a matter of fact, these conditions *are not fulfilled* in a particular instance. Still, do they imply that applying 'is a prime number' to Caesar is meaningless? Well, yes, in so far as understanding (3) requires understanding the operation of arithmetic division, and the latter applies only to numbers. And understanding (2) requires understanding that a Roman general is a flesh-and-blood person, not an abstract object like a number. That is also why Whiting's explanation of (1) fails. Given the

(1*) The famous Roman general is divisible only by 1 and himself

established meaning of the pertinent expressions, nothing counts as dividing a person—something that isn't a number—by 1 or by itself.

Knowledge of explanations and other semantic rules is neither necessary nor sufficient for saying something true. Whiting rejects this distinction between using an expression in a semantically correct way and using it to say something true. Wittgenstein demonstrably thought otherwise, from the *Tractatus* (4.024) through the distinction between the 'question of truth' and the 'question of meaning' (MS 106, 46) to his later insistence that the criteria both of linguistic understanding and of acceptable explanations of meaning are tied to conditions for the application of an expression rather than to knowing whether or not they are fulfilled (e.g. PI §§ 26, 49).

Turning to the bottom-up strategy, PPF § 309 and § 315 show that Wittgenstein continued to distinguish between sense and nonsense. They further imply that this is a matter of combination—a matter of what arguments are substituted in 'x knows what y is thinking'. According to Whiting, by contrast, the sentence 'It is correct to say "I know what you are thinking", and wrong to say "I know what I am thinking"' (PPF, § 315) does not give 'Wittgenstein's explanation of why those words make no sense in that combination. Rather, it represents an attempt to find a sense for the philosopher's words' (59). But that suggestion is remote from the text and incompatible with Wittgenstein's repeated insistence that, as intended by 'the philosopher', 'I know that I am in pain' is nonsensical.

Another passage I invoked is PI §§ 498–499:

> When I say that the orders 'Bring me sugar!' and 'Bring me milk!' have a sense, but not the combination 'Milk me sugar', this does not

> mean that the utterance of this combination of words has no effect. And if its effect is that the other person stares at me and gapes, I don't on that account call it an order to stare at me and gape, even if that was precisely the effect that I wanted to produce.
>
> To say 'This combination of words has no sense' excludes it from the sphere of language, and thereby bounds the domain of language.

Whiting avers that the reference to combinations of words may be just another way of referring to sentences. And 'there is no suggestion that those combinations lack sense *in virtue of* the meanings of the words they contain, the ways in which they are combined, or the rules governing their use' (60). But why should Wittgenstein all of a sudden speak of sentences as combinations of words when discussing the bounds of language? No monist answer is in sight. Furthermore, a reference to grammar and hence to rules of language immediately precedes this passage (PI §§ 496–497). Most importantly, the quote flatly contradicts Whiting's reading: It makes being nonsensical dependent *not* on lack of effect *but* on a combination that 'is excluded by', i.e. runs foul of the established use of its constituents.

Reply to Constantine Sandis

I am very grateful not just for Constantine's heartening comments on my work and (salad days) running prowess, but also for his insightful criticisms. At issue in the first instance is this: Does Wittgenstein's remark 'If a lion could talk (*sprechen*), we could not understand him' (PPR § 327) envisage a lion that utters sentences of a natural language like German? Or is it about a lion that produces signals of a feline language of complex growls, roars, etc., that we could never come to learn? Sandis' first piece of evidence for the first option is uncompelling (69). He maintains that we more readily allow that animals *talk* than that they *speak*, which indicates that, in using *sprechen,* Wittgenstein intimated human speech. I do not think that there is such a sharp contrast in English, and it is definitely absent from German. Right now, I am looking at an elementary introduction to animal communication entitled *Wie Tiere sprechen und wie wir sie besser verstehen*—'How animals speak and how we can better understand them'.

Of greater concern is Sandis' criticism of my argument in favour of the second option, namely that we would obviously be able to understand a lion uttering English sentences. Sandis rightly points out that I ignored the possibility that Wittgenstein was concerned with the problem of what sentences, phonetically belonging to a natural language, (could possibly) mean in the mouth of a lion. But his elaboration of that point is contentious, namely that if they meant what we mean 'we would not be dealing with a talking lion at all but with a human "in the shape of a lion"' (70). I explicitly allowed for the worry that it is debatable

whether 'such a talkative creature could count as a lion' (Glock 1996c, 166). That is an Austinian observation, however, not the one of PPF § 327. That passage takes at face value the possibility that a lion speaks. Nevertheless, Sandis raises a serious objection. By the lights of my own 'hermeneutic anthropology', there is not enough overlap in forms of life to understand an 'anglophone' lion. Sandis takes this obstacle to concern 'interpersonal understanding', whether it be of lions or completely alien humans. That is, indeed, the topic of PPF § 326, and I concur with Sandis and von Savigny that this context is relevant.

For better or worse, however, what I had in mind is the challenge posed by 'radical translation', which, in the first instance, is to make sense of what speakers of an entirely alien language mean by their specific utterances. And such *speaker occasional utterance meanin*g does feature in PPF § 325. Indeed, Sandis' own discussion revolves around a *semantic* problem, namely determining the concepts expressed by an anglophone lion. On that score, I want to stand my ground.

Consider a lion with highly advanced feline cognitive capacities who has just devoured an antelope, and lazily lies around licking its paws rather than attacking. Now moving to the fiction of the first reading, imagine that it also utters sentences like my example 'I'm not interested in you, I've just eaten an antelope'. In that case, I see no difficulty in principle. All of the concepts involved can be attributed to an intelligent feline, all the more so if its acoustic behaviour allows us to match parts of speech to appropriate features of the environment. If Wittgenstein's thought experiment is to raise a real issue, it is that of finding something like determinate content for the mental states and communicative behaviour of cognitively advanced yet non-linguistic animals. To solve that problem would require to transpose lessons from hermeneutic anthropology to hermeneutic ethology (see Glock 2020b). Some of the illuminating remarks that Wittgenstein and Sandis make about the former also shed light on the latter, or so I hope.

Reflections on Wolfgang Künne's contribution

I have embraced Wolfgang's modest or minimalist theory of truth at least since the eye-opening seminar he mentions in Endnote 5. It came to feature in my attempt to bring out the modest kernel of truth in the theories of Quine and Davidson. My only reservation was that Künne's magisterial *Conceptions of Truth* gives the impression that minimalism is tenable only if non-nominal quantification into sentence position can be reconciled with orthodox interpretations of quantifiers in formal logic. What matters in my view, however, is this:

> there is a perfectly good *colloquial* way of expressing the idea behind minimalism [. . .]. No matter what one says (believes, etc.) or

> could say (believe, etc.), what one says is true iff things are as one says they are. "Things" here does not indicate worldly items which make what we say true, it is simply the colloquial equivalent of a sentential variable.
>
> (Glock 2003a, 135)

Künne's current solution of 'Problem No. 1' seems to go in a similar direction. Non-nominal quantification into sentence position is both *sui generis* and legitimate, since it can be explained in non-technical terms (89).

The Wittgensteinian truism that Künne starts out with is a point in case. Although it informs the *Tractatus* as well, the official theory of that work is that a proposition is true iff the state of affairs it expresses obtains. This obtainment theory is of great interest not just because it provides the starting point for Wittgenstein's later account, but also because it can be glossed both as a kind of correspondence theory and as a minimalist theory (see Glock 2006). Doing so brings to the fore a point recognised by the early Wittgenstein and later emphasised by Ramsey: Once truth is elucidated in modest fashion, the challenge shifts to explaining the nature of truth-bearers and thence the topics of meaning and content.

Künne contributes very valiantly to this gargantuan project, by showing how the notion of a proposition can be used in a definition of truth without circularity. He acknowledges that 'it is a necessary a priori truth that propositions are "truth-evaluable", i.e. that they are things which can be evaluated in the dimension of truth and falsity', while resisting the conclusion that 'one cannot explain the meaning of "proposition" without recourse to the concept of truth' (90). True enough. But this resistance raises the more general question of how to distinguish definitional from non-definitional necessary truths. As I argued in my contribution to Künne's *Festschrift* (Glock 2011), the Wittgensteinian link of meaning to explanation and understanding is helpful here. The definitional truths must include *at least* those which feature in canonical explanations of an expression, and *at most* those acceptance of which is a precondition of understanding it.

A final ingenious and attractive move in Künne's piece is to allow that accepting propositions, i.e. sayables and thinkables, does not commit one to the existence of a category of entities that exist independently of anyone ever thinking or saying something. The example of Uncle Tom's cabin demonstrates the need to distinguish between individual and generic independence. Yet we still require an explanation of why a world without speakers or thinkers would also be a world without propositions. A possible solution might be that it is 'true about' a possible world without speakers that there are true propositions, but not 'true in' a possible world without speakers. Alas, my grasp of this modal

dichotomy is tenuous. For this reason, I rest content with a conceptual observation (see Glock 1997). Whether or not there are 'truths without people', the notion of 'a truth' makes sense only by reference to the idea of what rational subjects could say or think. My last two points are linked to a minor quibble. According to Künne (84), the 'really' in 'at least sometimes things (really) are as Galileo says they are' is merely a 'reading aid'. I regard it instead as a helpful reminder of a contrast that is partly constitutive of the concept of truth. While truths are sayables and thinkables, it is always possible that how things are said or thought to be differs from how things really are (self-evident truths aside).

Reply to Ansgar Beckermann

Ansgar's survey of contemporary philosophy of religion is concise and lucid. I agree with virtually all of his points. Some of them I have lodged myself. The principle of causality fuelling the cosmological argument is *neither* a scientific *nor* an analytic *nor* a *sui generis* metaphysical truth. Nevertheless, like the supervenience of macro- on microscopic properties, it is supported by a methodological principle that has guided scientific explanation since the seventeenth century. One should always look for a causal explanation, yet it is conceivable that no such explanation is forthcoming for reasons of principle, notably at the quantum level (Glock 1996a, 72–76; 2003b, 41–42).

By contrast, Beckermann's characterisation of my metaphilosophical stance is inaccurate. And his diagnosis that conceptual analysis plays at best a minor role in the philosophy of religion is unwarranted. First, the piece that Beckermann refers to explicitly denies that conceptual analysis constitutes the whole even of theoretical philosophy (Glock 2017b, 100; see also Glock 2008a, 99–105). I characterise the vocation of theoretical philosophy as 'critical thinking writ large' (Glock 2017b, 83). Beckermann's essay is a beautiful contribution to this genre. It does not engage in empirical research. As he himself emphasises, empirical findings from other disciplines do not even go a long way towards solving many of the problems it addresses. That requires a mix of methods, which includes considerations concerning the plausibility-cum-rationality of holding certain premises. Beckermann is right to point out that some of these considerations are difficult to classify, yet wrong to suggest that I cannot accommodate this. Thus, I maintain that there are methodological issues which straddle the distinction between conceptual and factual (Glock 2017b, 104–106).

Secondly, while assessing the credibility of premises is essential to thinking critically about religion, or any other topic, it is equally essential to establish 'what can be concluded from what' (118). Deductive inferences play a crucial, though not exclusive, role here (see 110). However, their validity *depends on the concepts* involved in premises

and conclusions, logical concepts in the case of formal inferences, additional nonlogical concepts in the case of material inferences. Beckermann himself rightly relies on a certain understanding of 'nothing' and 'exists' (103, 106). He suggests that such an understanding is only 'perhaps' a matter of conceptual analysis. In that case, however, and as a matter of propaedeutic conceptual clarification, he owes us an explanation of what he means by 'conceptual analysis'.

Thirdly, in line with contemporary 'conceptual engineering', Beckermann claims that even reflection about the *concepts* of religion and of God is 'not primarily about analysing how these concepts are used but rather about what is the best way to understand these concepts'. This is at odds with the moral he sapiently draws from his investigation of different ways of employing 'God': 'it does not seem appropriate to distinguish between 'right' and 'wrong' concepts of God' (107–108). Furthermore, assessing and improving concepts in a controlled and fruitful way presupposes at least a minimal understanding of the conceptual *status quo*; this is itself a conceptual point about methodology.

Fourthly, in promoting 'impure conceptual analysis', I intended to draw a deliberate contrast to orthodox conceptual analysis. Nevertheless, Beckermann writes:

> In my view, Glock holds that the mentioned factual questions can be answered only if first the corresponding conceptual questions have been answered. This is the job of philosophy. After that is done the questions are handed over to the empirical sciences.
>
> (101)

However, I stress that while conceptual and factual issues can be distinguished, they typically interact and that this overthrows 'a received image of conceptual analysis, namely that it is a purely *a priori* exercise, unaffected in all respects by scientific findings' (Glock 2017b, 100). In the same vein, I diagnose various 'mutual dependencies of conceptual and factual considerations' (Glock 2017b, 103). I also warn against the 'Socratic mistake of thinking that one cannot establish empirical facts about X unless one already has an analytic definition' of 'X'; and I contend that providing a definition 'marks at best the terminus of philosophical inquiry, not its beginning' (Glock 2017b, 101). This is the opposite of the position that Beckermann ascribes to me. What I insist on is that any investigation, philosophical or scientific, relies on a certain *preconception* of its topic, and that this preconception may always come to require explication and clarification. Beckermann's disquisitions about religion do not cast doubt on these conceptual truths. Instead, his procedure willy-nilly supports my contention that conceptual clarification remains an 'indispensable tool of critical thinking' (Glock 2017b, 96).

Reply to Christian Nimtz

Christian inveighs against what he describes as 'Glock's key axiom that philosophy is essentially concerned with conceptual matters, if that claim is to be more than a triviality perfectly amenable to any philosophical naturalist averse to analytic truths'. I do, indeed, believe that the distinctive contribution philosophy has to make *includes* the clarification of conceptual issues as its central component. But that presupposes, of course, that there are *distinctively conceptual* issues and connections, an idea which is intimately related to analytic truths. Plenty of otherwise diverse naturalists dispute not just the idea of analytic truths but the very existence of concepts (aka 'meanings'), and they abjure *a priori* reflection (e.g. Quine, the Churchlands, Kornblith, Ladyman). As an impure conceptual analyst, I am happy to agree with those self-confessed naturalists who assign a central philosophical role to conceptual reflection and accept that empirical findings of science can be relevant to philosophical issues, yet without settling them. Nimtz does not dispute this stance. Instead, he argues that there is also the opposite direction of input. Some armchair reflections yield '*de facto scientific hypotheses*' (122). In that respect, his paper might equally well be entitled: 'Some Good (Scientific) News *from* the Philosophical Armchair'.

Nimtz's case rests on imaginary scenarios. However, if such scenarios are to support (proto-)scientific abductive reasoning rather than mere conceivability in near-actual worlds, they would have to be not just conceptually but also physically possible. Now, some physical possibilities are so commonplace and basic that they can be taken for granted in armchair reflection. By contrast, others can only be ascertained through scientific observation, experimentation, and the ensuing theories. In those cases, the news from the armchair will rely at least partly on input from empirical science.

Nimtz's main exhibit is the 'Indexicality Lemma' that he skilfully extracts from Perry's puzzle-cases (124). Rephrased so as to bring out connections to contemporary theory of action, it runs:

IL Any Perry-type intentional action [any intentional action *A* performed by a subject *S* that occurs because of a change in *S*'s beliefs about where *S* is located in space and time] is such that *S*'s reason for *A* must comprise a truly-indexical belief [a belief with a non-eliminable indexical content such as 'I am [. . .] ,' or 'This is [. . .] '].

According to Nimtz, IL is a '*de facto scientific hypothesis*', on account of three features.

a IL is *a posteriori*.
b IL is metaphysically contingent.

c IL 'falls squarely into the purview of the empirical sciences of the mind' (125).

(a) is correct only in so far as the scenarios from which IL is derived and to which it applies are of a mundane kind frequently encountered by normal humans. As such, they fall within the purview 'of the commonplaces we are familiar with anyway' (121), something that both Nimtz and I reckon with. At the same time, the derivation of IL from these scenarios may be fallible, yet it is not empirical in any recognisable sense.

In opposition to (b), my impure conceptual analysis explicitly rejects the idea of *sui generis* metaphysical modalities. In any event, IL is best seen as a non-obvious conceptual truth. According to a non-psychologistic analysis of reasons for action propounded by five symposiasts (Alvarez, Hyman, Schmidt, Steward, and myself), the reason for an intentional action is a fact or presumed fact that makes the action attractive in the agent's eyes. From this non-psychologistic analysis in conjunction with Nimtz's definitions of 'Perry-type action' and 'truly indexical belief', it follows that the reason for a Perry-type action must include a belief referring to the agent's own location in space and time. The reason for the shopper rearranging the torn sack in his cart is the fact that *his* shopping cart is making a mess. It is this fact that makes rearranging the sack attractive to the shopper, not facts concerning somebody else's shopping cart.

By a similar token, contrary to (c), IL belongs to the empirical sciences of the mind only in so far as their conceptual framework includes notions like intentional action, reason, and motive. That IL sustains counterfactual conditionals is correct, yet does not set it apart from claims that Nimtz would accept as conceptual. Take an example close to the hearts and minds of other symposiasts. If it were *true* that the last presidential election in the United States was fraudulent, then this is *how things would be*: The last presidential election in the United States was fraudulent. The only empirical input to Nimtz's procedure is the fact that we often respond differently to a situation once we realise our spatial and temporal relation to it. What Nimtz contributes is a reconceptualisation of that fact. He thereby helps to *clarify* an *explanandum* of cognitive science. That is *not* the same as doing empirical research; yet it is good news from the armchair nonetheless!

Reflections on John Hyman

I have been greatly privileged to learn from John's seminal writings, and I can hardly intimate how much I have benefited from discussing philosophy with him over so many years. Part of this privilege has been following the development of his thought on perception.

There is little to add to Hyman's compelling critique of the modern causal theory and the empiricist assumptions it shares with disjunctivism, except for two remarks. Prodded by Hyman's homework assignment to detect the fallacy in the simple-minded 'BBQ argument' for the causal theory (138 and n. 5), I should like to distinguish two claims concerning the consequences of

1 A smells X (lamb chops) by smelling Y (emitted vapour) ⇒ X and Y are causally related.

Namely

2 ¬ ((1) ⇒ (A smells X by smelling Y ⇒ *smelling* X and *smelling* Y are causally related))

and

3 ¬ (A smells X by smelling Y ⇒ *smelling* X and *smelling* Y are causally related)

2 is definitely correct. But to move from (2) directly to (3) would beg the question against the causal theory, since the latter maintains that there is a conceptual connection between the left- and the right-hand side of the arrow.

My second remark concerns Hyman's insistence that 'the ordinary concept of perception' does not rule out unconscious perception. 'Otherwise [unconscious perception] would not be a discovery, it would be a contradiction in terms' (137). Given their distinction between altering theories and modifying concepts (which both Hyman and I endorse), Grice and Strawson could respond that accepting unconscious perception amounts to a *modification* of the ordinary concept, albeit one undertaken in the light of empirical discoveries. Accordingly, more needs to be said on this score.

Hyman argues that the ordinary concept of perception signifies a 'multi-track disposition', a multifaceted and plastic responsiveness to information provided by the environment (156–157). This avoids not just the empiricist myth of perception as a 'seeming-to-perceive with a suitable cause'. It also militates against what I have called the rationalist 'myth of spontaneity', according to which genuine perception is a capacity to (self-)consciously, actively, and intentionally apply concepts in perceptual judgements (Glock 2021). In a higher animal, this multi-track disposition is closely connected not just to a capacity for pursuing goals, but also to the control of its own sense-organs (157). Here I detect an affinity to my suggestion that our notions of behaviour and perception constitute a conceptual *Gestalt* which intertwines behavioural with morphological-*cum*-anatomical features. Perception requires not just a certain reaction to information provided by the environment, but also

sense-organs dedicated to gathering such information. And the paradigm of animal behaviour is goal-directed activity along a line of orientation determined by the position and alignment of these organs.[2] That is one reason, I submit, why Hyman's *Überwindung* of empiricism provides an auspicious framework for integrating the geometric, physical, physiological, and affective aspects of perception.

Reply to Markus Wild

This is more than a *little* help from a friend. The distinction between hermeneutic and hydraulic ethology is a great foil for capturing the differences between Markus and me. In so far as the two conflict, it is with respect to three issues connected to what he calls my 'pure capacity approach'.

First, Wild doubts that an appeal to capacities could 'constitute good philosophy' while leading to 'bad science', since 'empty or circular explanations' (170) are no use in either field. He does not address, however, my contention that we are dealing with explanations of different kinds: *constitutive/conceptual* on the one hand, *causal/genetic* on the other. The capacity approach non-vacuously explains what it is for a subject *S* to possess a mind not as *S* being a composite of a body and a mental substance or as *S* undergoing neurophysiological processes, but as having a suite of abilities like sentience, perception, conation, emotion, reason, will, etc.

To that extent, my pure capacity approach is compatible with Wild's teleosemantic capacity approach. One point setting them apart is that the latter characterises all mental powers as capacities for 'having representations'. My reservations about this monistic perspective are twofold. First, some mental states such as mere sentience or moods cannot be squeezed into the Procrustean bed of representing an object or content. Secondly, regarding intentional states, the notion of representation is *either* too specific, e.g. when it is spelled out in pictorial or symbolic terms—*or* it boils down to the claim that these states have reference/accuracy/truth/satisfaction conditions. While that claim is unobjectionable, it states an *explanandum* rather than an *explanans*, for both conceptual and causal explanations.

This is where a final contrast arises. Teleosemanticists pursue constitutive explanations (many, though not Wild, would shudder to call them conceptual) through a causal explanation concerning the genesis of these states and capacities. While the pure capacity approach acknowledges the enormous scientific relevance of genetic considerations, it denies that a particular origin is conceptually necessary or sufficient for intentional states. I may be prejudiced in favour of Wild's delightful Swamp-Swabian—after all, he can run, write philosophy articles, and make *Spätzle*! But most competent speakers would concur with me that he possesses cognitive and conative capacities that qualify as mental.

Wild rightly points out that there are some intentional predicates that Swamp-Swabian cannot satisfy. He cannot *re*-cognise a perfect dough or remember the recipe. This shows that *some* mental predicates have a genetic dimension which is interestingly related to (though not identical with) the factive aspect of 'knows', 'recognises', etc. Nevertheless, as long as one employs these terms in the way established in both everyday life and behavioural science, it seems clear that Swamp-Swabian can see and thereby correctly judge that a *Spätzle* dough has the right consistence. Accordingly, a blanket genetic account of all intentional vocabulary appears unwarranted. Note, finally, that the intellectual and volitional capacities of Swamp-Swabian will enable him immediately to acquire some 'genetic properties'.

Reply to Maria Alvarez

The distinction between justifying, motivating, and normative reasons is a kind of holy trichotomy of contemporary theory of action. No one has done more than Maria to explain this trichotomy and to demonstrate its potential for clarifying the nature of practical reasons. My contrast between objectivist and subjectivist conceptions was explicitly meant to apply to motivating reasons, the reasons *for* or *in the light of which A Φ*-s. It is the question of whether animals can act for *motivating* reasons that Alvarez and I disagree on. I continue to think that the official line of the Davidsonian mainstream regards motivating reasons as mental states of believing, desiring, etc., or their onslaughts. This commits them to the view that acting in the light of reasons means acting in the light of one's own mental states. I am happy to concede, however, that if subjectivism is restricted to *explanatory* reasons, it will not imply that acting for reasons entails self-ascription of mental states.

Alvarez is dead right that the crux of our controversy is whether agents can be in states of believing and desiring whose *contents* can guide them in acting. Barring the possibility that goals can be motivating reasons of an objective kind (as argued by Glock and Schmidt 2021), this implies that the question of whether animals can act for motivating reasons now turns on two issues: First, can they believe or know facts or states of affairs? Secondly, can these facts favour doing something in their eyes?

As regards the second question, Alvarez and I concur that intelligent animals are capable of evaluating things and of acting on such evaluations. What divides us is the question of whether they evaluate facts or something less that-ish, like Gibson's affordances. The 'ontological' nature of affordances has turned into a hot topic of debate. In the current context, I should merely like to insist that they do not afford an obvious and uncontentious bypass of that-ishness. Alvarez concedes my claim that animals can cognitively respond to 'features of their environment [. . .] as good or bad (e.g. attractive or threatening) [. . .] and act accordingly'

(198). However, at the risk of being an intellectual hedgehog, the question remains whether and in what respects cognising (recognising, perceiving, etc.) *a as F* falls short of cognising *that a is F*. Let me finally note that for all my resistance to lingualism, I am in full agreement with Alvarez in holding that the ability to deliberate and reflect on reasons *qua reasons* is *both* tied to language *and* a decisive game-changer when it comes to the role that acting for a reason plays in the life-form of a species, a game-changer distinct from, yet not independent of, the unduly neglected phenomenon of intelligence.

Intermezzo: A dog's breakfast?

I cannot move on from the contributions of Wild and Alvarez without paying a visit to a certain dining room. In my view, animals can act in the light of reasons because they can act in the light of facts they perceive. In support, I introduced the case of a dog that has learned not to grab an object when it is lying on the table but only when the object is lying in his bowl. This dog sees a bone on the table at t_1, yet refrains from grabbing it until, at t_2, it has been placed in the bowl. I contended that this feat cannot be explained by the dog perceiving objects or conglomerations of objects, but only in terms of the following opposition:

(I) The dog sees at time t_1 *that* the bone is on the table.
The dog sees at time t_2 *that* the bone is in the bowl.

Both Wild and Alvarez find this argument unconvincing. With apologies to the former editor of the *Guardian* C.P. Scott, their stance concerning the contents of perception is epitomised by the slogan: 'Objects are free, but facts are sacred!' They suggest a variety of potential contents for the dog's perceivings that are non-that-ish, but which my considerations do not rule out as *explanantia* of the dog's behaviour (unlike Alvarez, Wild holds that they *can* be ruled out by 'hydraulic' findings). I can only comment on two of these alternatives.

According to Wild, the dog might perceive two distinct perceptual objects—'tabley-bone' and 'bowly-bone'. But this does not defang my line of reasoning. We are, in fact, dealing with a single object rather than two distinct objects. Furthermore, dogs can *recognise* this. They satisfy standard tests for object permanence and identification (Miller et al. 2009), though less comfortably than apes (Seed and Tomasello 2010, 409). And there is no question that our dog does not just perceive two distinct perceptual objects, but visually *tracks* an *enduring particular* on its path from table to bowl between t_1 and t_2. Wild conjectures that the dog might perceive a whole series of distinct perceptual objects between t_1 and t_2 (174, n. 8). Alas, that attempt to minimise the *types of cognitive capacity* with which the dog is credited leads to an unacceptable

proliferation of the *number of cognitive operations*. Such a rapid series of perceived objects would send my mind (and that of Swamp-Swabian) into a spin, not to mention the mind of a humble dog.

According to Alvarez, at t_1 and t_2 the dog perceives 'different arrangements of objects'—bone on table vs. bone in bowl. And such an arrangement is not 'some abstract thing over and above the objects so arranged' (196). Now ask yourself: Arranged how? Well, such that the bone is on the table at t_1 and in the bowl at t_2. That is the most natural answer, just as (I) is the most plausible explanation of the dog's behaviour. Admittedly, there is an alternative response: 'Like this ☞' pointing at two different situations. But the challenge remains of specifying these situations in a way that satisfies two conditions: First, it avoids what I call the 'behaviourist fairy tale' of animals experiencing mere stimuli, secondly, it does not lapse into that-ish idiom. Unless the components of the situation are perceived as arranged in such-and-such a way, we are back with Wild's suggestion of a mere difference in (admittedly complex) objects. To conclude: On the one hand, Alvarez's attributive 'so arranged' is essential to her explanation of the dog's demeanour, since unlike the arranged objects or their sum, it sets the two situations apart; on the other hand, that attributive must allow of being spelled out that-ishly.

Reflections on the contribution by Albert Newen, Maja Griem, and Simone Pika

Newen, Griem, and Pika propose a new conceptual framework for studying empathy. They first identify central components in paradigmatic human cases, then distinguish types of empathy emerging in human ontogenesis, and finally consider which of these types are present in non-human animals according to current ethology. This is the most promising procedure for bringing conceptual elucidation and precisification to bear on cognitive science, in the constructive way which has become a trademark of the 'Bochum School' founded by Albert. I also sympathise with the family resemblance account of how these central components make for different notions of empathy, and with their avoidance of the over-intellectualisation inherent in the popular view that all empathy requires imagining the mental states of others.

The authors discern three strands within the family of empathetic phenomena—(1) registration of someone else's mental state, (2) caring attitude, and (3) supportive behavioural response (206). They duly note that in the case of non-linguistic subjects, it is difficult to hold (2) and (3) apart. How can one establish that a subject has a caring attitude even when it lacks any opportunity for helping or consolation behaviour? In my view, a possible solution takes recourse to physiological measures and facial expressions, as well as to cognitive distortion and cognitive biases (Burman et al. 2009). But if the difficulty cannot be overcome, this

might also indicate that (1)–(3) emerge as *clearly distinct* components of empathy only as the result of a diversification which requires the acquisition of language, both at the phylogenetic and at the ontogenetic level.

Finally, although it is not directly implied by (3), the paradigmatic examples that Newen, Griem, and Pika operate with and their emphasis on helping and consolation behaviour suggest that empathy is primarily a matter of feeling sorry. But one can also care for—and empathise with—others in a less defensive or remedial spirit, for example, by taking delight in their successes. In that case, it is at least tempting to explain component (2) as *feeling with others*. To do so, however, would introduce a degree of circularity to the explanation, notwithstanding the fact that the distinction of different types and levels of empathy provided by the authors is enormously informative.

Reply to Helen Steward

It is gratifying to note that such an eminent theorist of action agrees with much of my case against 'inflationist' conceptions of agency. And it is enlightening to face pressure, for once, from the opposite direction, that of a 'deflationist' conception. Helen accepts my 'perception-behaviour Gestalt' claim in so far as our 'inclination' to regard an organism as capable of perception is entangled with our inclination to regard it as capable of behaviour. At the same time, she argues that *both* perception *and* behaviour may extend to plants and some micro-organisms, since these inclinations may be part of a 'deeply problematic and partially hard-wired zoocentrism' (230). Now, I plead guilty to a *conceptual* zoo- or even anthropocentrism. Our concepts of perception and behaviour play a role in *our* cognition, serve *our* epistemic needs and interests, and are geared to *our* capacities. To that extent, they are anthropocentric; yet they are none the worse for that! It does not follow that it is anthropocentric to insist that these concepts preclude application to organisms bereft of locomotion and sense-organs. Still, should these concepts not be replaced by less parochial alternatives, e.g. those of 'information integration theory' (231–233)? Leaving details aside, my challenge to such a profound revisionism is to provide a principled stop to a slide into panpsychism (see Glock 2020b).

Like Steward, I am keen to keep the question of what separates animals from plants apart from the question of how to distinguish different types of activity, this being the most general category I recognise (Glock 2019, 648–649). She is right to complain, however, that my demarcations remained unclear (234–237). Here is an attempt to make amends. Mere *animate activity* is determined by needs and guided by information-processing. *Behaviour* is goal-directed and informed by perception. *Agency* requires a differentiation between cognitive and conative determinants of behaviour, and hence a capacity to act on 'wants',

goals that the subject can adopt. At this point, my article muddies the waters by an excursion into intelligence and rationality and by engaging with debates about the relation between action, intentional action, and acting for a reason. Nevertheless, my position is coherent. *If* intentional action is conceived as acting in pursuit of a goal of one's own (as I would suggest), then all agency in my sense (pursuit of goals guided by perception) is intentional. *If* intentional action equates to acting for a reason, then *either* goals are reasons (as Glock and Schmidt 2021 vouchsafe), in which case once again all agency is intentional (Glock 2019, 657–660), *or* intentional agency is confined to responsiveness to reasons / facts. It is only this more demanding conception of intentional agency which requires the kinds of that-ish cognitive and conative states that Steward, Wild, and I are content to attribute to intelligent animals—by contrast to Alvarez and Ben-Yami (both this volume).

Accordingly, the leopard frog is a basic behaver, and an animal pursuing diverse goals yet incapable of learning to distinguish facts relevant to those goals (unlike the dog taken for an intellectual outing by Wild and Alvarez, both this volume) is a basic agent. I fully agree with Steward that one also needs to draw distinctions on the basis of the stimulus-response decoupling she favours. But the latter is close to intelligence, a purely cognitive notion that I decided to keep separate. Most importantly, I am in full sympathy with her quest for achieving a stable middle-ground between inflationism and deflationism. As I see it, this confronts us with a *seesaw effect* that afflicts the philosophy of animal minds more generally. On the one hand, conceding relatively undemanding mental powers to non-human animals seems to commit one to granting many or even all advanced capacities. On the other hand, denying animals highly sophisticated mental powers seems to force one to deny them even very simple capacities. Steward and I are positioned on different sides of this seesaw, yet we both try hard to stay close to its linchpin.

Reply to Hanoch Ben-Yami

Hanoch and I agree on one important issue. Mastery of basic logical concepts like *disjunction* and *negation* does not *conceptually* depend on language. I have argued for this explicitly (e.g. Glock 2017a); he implicitly acknowledges it by maintaining that it 'is an empirical question' (244) whether (non-linguistic) animals possess such concepts. As regards scientific findings purporting to demonstrate advanced mental capacities in animals, Ben-Yami has for some time played the role of a tenacious scoffer. His doubts are always noteworthy and often warranted. I fully agree with his verdict on diagnoses of animal morality (248), though on different grounds. Even if some monkeys display *egocentric* inequity aversion, the latter falls short of *allocentric* inequity aversion and thereby of *bona fide* morality (Christen and Glock 2012).

In other cases, Ben-Yami's scepticism strikes me as excessive. He contests that the landmark experiment of Clayton and Dickinson shows that the cache recovery pattern of scrub jays fulfils the three 'what', 'where', and 'when' criteria for episodic recall (see 250). According to him, it shows 'that scrub jays expect to find food of the kind cached where they cached it, not that they remember the episode of caching it there; and that their motivation to recover perishable food declines with time' (250). But the experiment indisputably demonstrates that the recovery patterns concerning worms and seeds differ along the *what-where-when cached parameters*. And it remains a mystery why the jays should expect to find food of a certain kind where they cached it, without remembering caching it. As regards Ben-Yami's alternative explanation, it is excluded by the control group. Jays that had *not* learnt that worms go bad did not show the recovery pattern of those that had. There is no indication whatsoever that the motivational state of the latter changed. All jays continued to prefer worms over seeds, but this preference was overridden by one for preserved over spoilt food. This is the only explanation of their recovery pattern; and it is, in turn, best explained by granting that they remember when and where they cached what type of food.

Even if Ben-Yami had dismantled the Clayton/Dickinson case for episodic memory, it would remain a fallacy to proceed, as he does, to the declaration that it is a 'fact that animals do not have episodic memory' (250). This is an instance in which Ben-Yami moves illicitly from objections to cases for the presence of certain cognitive capacities in animals to conclusions about their absence. Let us grant for the sake of argument that all the studies suggesting disjunctive reasoning (including ones that Ben-Yami (245–246) does not mention, such as Call 2004) are at best inconclusive. While a capacity for disjunctive reasoning entails a grasp of negation, the reverse does not hold. Although Ben-Yami seems to recognise this point, he goes on to assume that no animal has a grasp of propositional negation. And on this shaky ground he proceeds to asserting that no animal has an understanding of possibility, without confronting numerous studies suggesting otherwise (248). To mention just one, a study by Mulcahy and Call (2006) indicates that great apes can learn to reckon with the *possibility* that tools hitherto available may not be so in the future.

Finally, Ben-Yami's brief comments on sign-use in animals run together two issues. One is whether the *sign-communicative function relation* is arbitrary in the sense that the community of sign-users have the option of altering it, for example, by using new signs to serve an established function. The other is whether individual users can have *communicative intentions* in using a sign on specific occasions.[3] Even if many native Anglophones were unable to switch to sign-function mappings other than those familiar from English—as an uncharitable interpretation

of anecdotal evidence indicates—this would not entail that they lack 'communicative intention' (246): They use English expressions with the intention to communicate, and the way they use these expressions is for the most part subject to their control.

Reflections on Julia Langkau's contribution

Julia starts out from Boden's influential definition of creativity as the ability to come up with something new, surprising, and valuable. She shows that this paradigm has yielded verdicts that conflict with our pre-theoretical intuitions over ascribing creativity to a) non-human animals; b) AI systems; c) outcomes lacking value. She wisely counsels against shrugging off this difficulty simply on the grounds that Boden's stance is coherent. Instead, she diagnoses a 'conceptual tension' to be resolved by clearly distinguishing two notions of creativity run together in current debates: product creativity and process creativity. To clear the ground, Langkau resists two tempting claims, namely that creativity should be defined or understood by reference to either the reactions of others or the subjective feelings of the subject. She goes on to complain that creativity should not be understood in terms of the *abilities* of a *subject*, on the grounds that philosophers should be interested in underlying mental phenomena rather than outward behaviour. An ability is not an efficient cause of its exercise, to be sure. But neither is it identical with that exercise. Indeed, one can invoke abilities in favour of Langkau's approach. An ability is individuated by reference to its exercise (see Glock 2010). And this implies that in explaining creativity we must decide whether the constitutive exercises of various *specific* creative capacities are to issue in a certain product or to instigate a certain process.

At this point, I have a second qualm. While I accept the importance of the product/process dichotomy, I wonder whether it is one between different *concepts* of creativity. Are 'creative' and its cognates genuinely ambiguous? Or is it rather a matter of two kinds of things that can be the contents of a unified phenomenon? After all, 'anger' is not ambiguous simply because its grammatical objects range over things as divergent as acts, omissions, character traits, people, institutions, and statements. From this perspective, process creativity could be subject to the very same parameters as product creativity. Regarding Boden's types of creativity: The process of producing an object can be novel simply in combining existing procedures, in expanding and modifying techniques, or in transforming the range of conceivable practices in an entirely unheard-of way. Conceiving of the difference between product and process creativity in this more modest way is in the spirit of Langkau's enterprise, and it suffices to resolve the unclarities and inconsistencies that she acutely diagnoses.

Reply to Brad Hooker

In so far as Brad's gracious compliments have any foundation *in rebus*, it is that I had the privilege of contributing to a golden age of philosophising at Reading, along with immensely accomplished colleagues like him. Hooker agrees with my distinction of four conceptions of rationality. I, in turn, welcome his insistence that one has to distinguish possession of the capacity of rationality from its successful exercise (276; see Glock 2019, 656). Hooker's authoritative survey of how rationality, reasons, and rules are connected in contemporary normative ethics certainly counts among the latter. He alerts us to the fact that we must further distinguish the capacity from the *requirement* to be rational (279). My four conceptions specify what constitutes being rational. His discussion opens up the question of *why one should be rational* in any of these senses. Hooker shows that there is no intrinsic general requirement to maximise one's own interest or to justify oneself *vis-à-vis* others. In the sequel, he concentrates on the two remaining conceptions—being responsive to reasons and being capable of reasoning. To me, Hooker's most illuminating contribution lies in combining the idea that the latter is a subclass of the former (280) with acknowledging that rules cannot be eliminated in favour of reasons, at least in the moral realm.

He mentions four rationales, which all have an echo in my ruminations about normativity (285–288). First, rules are essential to classification and thereby to thought in general, moral thought included. Secondly, while there is always the question whether actually shared rules (especially of a moral kind) ought to be shared, this does not detract from the fact that those rules which mark generally accepted requirements generate reasons for action. Rules provide intrinsic reasons while at the same time standing in need for being justified by reasons (see Glock and Schmidt 2021). Thirdly, permissive rules allowing one to Φ do not generate direct reasons for Φ-ing; yet they are a prerequisite for acting freely according to one's reasons for or against Φ-ing. Fourthly, constitutive rules are essential to social practices that confer normative powers, including those of a moral type. In addition to Hooker's example of promising, accusing, excusing, and apologising immediately spring to mind.

Finally, without rules, we would lack the common ground that enables moral deliberation on specific issues. As I should like to put it: Rules liberate, by allowing us to go about our daily business, but also by facilitating rational reflection on what rules to adopt. By the same token, moral reasons may be irreducible to rules, yet they can make a difference to the lives of cooperative beings with limited cognitive and moral powers only through the scaffold of norms, even though this scaffold is inevitably loose and always subject to alteration.[4]

Reflections on Gerhard Ernst's contribution

Gerhard's excursion into the mind of an examinee is both entertaining and thought-provoking. He uses his seven cases to very good effect in arguing for a particular view of the rationality of emotions. Still, from a perspective without theoretical stakes, I am inclined to register more differentiated verdicts. Case 1 is clearly irrational and may even border on being unintelligible. Of course, John may still feel a residual anxiety, yet if he genuinely believes that there will be no exam it is unclear how he can also genuinely (never mind rationally) fear it. Case 2 is certainly intelligible and far less irrational, since being prepared is no guarantee of success. Case 3 is perfectly rational, unless one believes, mistakenly, that death is the only thing to fear. Case 4 is irrational from a cognitive perspective yet perfectly intelligible as an emotional reaction—hysteria manifested in hilarity. Case 5 is intelligible and, in my experience, not irrational. One can look forward to a real challenge while nevertheless feeling trepidations; indeed, it is possible to enjoy that very state of emotional ambivalence. At any rate, that was my experience on the mornings of running a marathon.

Finally, while there are palpable advantages of dealing with Ernst's exhibit cases by construing emotions as dispositions, there is also a worry. If beliefs, intentions, and emotions are dispositions, and if manifesting a disposition is a constitutive criterion for having these beliefs, intentions, and emotions, this militates against diagnosing an inconsistency between a determinate belief/intention/emotion combination and certain other belief/intention/emotion combinations in the way Ernst does (298). If getting up at 7 and not entertaining thoughts like 'Another 5 minutes won't make a difference' is a criterion for believing that one ought to get up at 7, then a person who fails on both counts does not act in a way that is inconsistent with her beliefs; rather, it is unclear *what beliefs* she holds in the first instance. So it could not qualify as a straightforward case of *akrasia*.

Notes

1 Unless otherwise indicated, page references are to this volume.
2 In so far as recognising morphological and anatomical as well as behavioural features in the behaviour/perception *Gestalt* is a concession to Wild's 'hydraulic ethology', it is one I am prepared to make. See Glock 2019, 649; 2020a, 217–8.
3 For this and related distinctions concerning sign-systems, see Gasparri et al. (forthcoming). For evidence of limits to arbitrariness in human languages and the presence of some arbitrariness in animal communication, see Watson et al. (forthcoming).
4 Let me mention a final inspiring thought in the paper: The analogy between confining maximisation of interests to a particular person (oneself) and confining it to particular places and times, while not furnishing a knock-down

argument, strongly supports altruism for those responsive to reasons. Just like the demands of individual agency apply equally to times and places, they also apply equally to each and every individual agent!

References

Burman, O.H.P., Parker, R.M.A., Paul, E.S. and Mendl, M.T. (2009). Anxiety-Induced Cognitive Bias in Non-Human Animals. *Physiology and Behavior* 98(3): 345–350.

Call, J. (2004). Inferences about the Location of Food in Great Apes. *Journal of Comparative Psychology* 115: 159–171.

Christen, M. and Glock, H.J. (2012). The (Limited) Space for Justice in Social Animals. *Social Justice Research* 25: 298–326.

Gasparri, L., Filippi, P., Wild, M. and Glock, H.J. (forthcoming). Notions of Arbitrariness. *Mind and Language*, 2022.

Glock, H.J. (1996a). *A Wittgenstein Dictionary*. Oxford: Blackwell.

Glock, H.J. (1996b). Abusing Use. *Dialectica* 50: 205–223.

Glock, H.J. (1996c). On Safari with Wittgenstein, Quine and Davidson. In Arrington, R. and Glock, H.J. (eds.), *Wittgenstein and Quine*. London: Routledge, pp. 144–173.

Glock, H.J. (1997). Truth without People? *Philosophy* 72: 85–104.

Glock, H.J. (2003a). *Quine and Davidson on Language, Thought and Reality*. Cambridge: Cambridge University Press.

Glock, H.J. (2003b). Strawson and Analytic Kantianism. In Glock, H.J. (ed.), *Strawson and Kant*. Oxford: Clarendon Press, pp. 15–42.

Glock, H.J. (2004). Was Wittgenstein an Analytic Philosopher? *Metaphilosophy* 35: 419–444.

Glock, H.J. (2005). Wittgenstein and History. In Pichler, A. and Säätelä, S. (eds.), *Wittgenstein: The Philosopher and His Works*. Bergen: Wittgenstein Archives at the University of Bergen, pp. 177–204.

Glock, H.J. (2006). Truth in the Tractatus. *Synthese* 148: 345–368.

Glock, H.J. (2008a). Analytic Philosophy: Wittgenstein and After. In Moran, D. (ed.), *A Companion to Twentieth-Century Philosophy*. London: Routledge, pp. 76–117.

Glock, H.J. (2008b). *What Is Analytic Philosophy?*. Cambridge: Cambridge University Press.

Glock, H.J. (2010). Concepts, Abilities and Propositions. *Grazer Philosophische Studien* 81: 115–136.

Glock, H.J. (2011). A Cognitivist Approach to Concepts. *Grazer Philosophische Studien* 82: 111–143.

Glock, H.J. (2015). Nonsense Made Intelligible. *Erkenntnis* 80: 111–136.

Glock, H.J. (2017a). Animal Belief and Rationality. In Andrews, K. and Beck, J. (eds.), *The Routledge Handbook of Philosophy and Animal Minds*. London: Routledge, pp. 89–99.

Glock, H.J. (2017b). Impure Conceptual Analysis. In Overgaard, S. and d'Oro, G. (eds.), *The Cambridge Companion to Philosophical Methodology*. Cambridge: Cambridge University Press, pp. 83–107.

Glock, H.J. (2017c). Philosophy and Philosophical Method. In Glock, H.J. and Hyman, J. (eds.), *The Blackwell Companion to Wittgenstein*. Marsden: Wiley, pp. 231–251.

Glock, H.J. (2019). Agency, Intelligence and Reasons in Animals. *Philosophy* 94: 645–671. https://doi.org/10.1017/S0031819119000275.

Glock, H.J. (2020a). Animal Consciousness – A Limit of Language? In Appelquist, H. (ed.), *Wittgenstein and the Limits of Language*. Abingdon: Routledge, pp. 197–222.

Glock, H.J. (2020b). Minds, Brains and Capacities: Situated Cognition and Neo-Aristotelianism. *Frontiers in Psychology* 11: 1–14. https://doi.org/10.3389/fpsyg.2020.566385.

Glock, H.J. (2021). Concepts and Experience: A Non-Representationalist Approach. In Demmerling, C. and Schröder, D. (eds.), *Concepts in Thought, Action, and Emotion*. Abingdon: Routledge, pp. 21–41.

Glock, H.J. and Schmidt, E. (2021). Pluralism about Practical Reasons and Reason Explanations. *Philosophical Explorations* 24(2): 119–136.

Miller, C.H., Rayburn-Reeves, R. and Zentall, T.R. (2009). What Do Dogs Know about Hidden Objects? *Behavioural Processes* 81(3): 439–446.

Mulcahy, N.J. and Call, J. (2006). Apes Save Tools for Future Use. *Science* 312: 1038–1040.

Schulte, J. (1990). *Chor und Gesetz*. Frankfurt: Suhrkamp.

Seed, A. and Tomasello, M. (2010). Primate Cognition. *Topics in Cognitive Science* 2: 407–419.

Watson, S.K., Filipp, P., Gasparri, L., Falk, N.G., Tamer, N., Widmer, P., Manser, M. and Glock, H.J. (2022). Optionality in Animal Communication: A Novel Framework for Examining the Evolution of Arbitrariness. *Biological Reviews*: 1–19. https://doi.org/10.1111/brv.12882.

Wittgenstein, L. (1961). *On Certainty* [=OC]. Edited by G.E.M. Anscombe and G.H. von Wright. Oxford: Blackwell.

Wittgenstein, L. (1967). *Zettel* [=Z]. Translated by G.E.M. Anscombe. Oxford: Blackwell.

Wittgenstein, L. (2009). *Philosophical Investigations* [=PI]; incl. *Philosophy of Psychology—A Fragment* [=PPF]. 4th edition. Revised by P.M.S. Hacker and J. Schulte. Oxford: Wiley Blackwell.

Hans-Johann Glock: A Bibliography

Monographs

A Wittgenstein Dictionary. Oxford: Blackwell (1996), x + 406 pp. [translations into Portuguese, German, Polish, French, and Turkish.]

La mente de los animales: problemas conceptuales. Oviedo: KRK Ediciones (2009), 174 pp. [Spanish monograph based on unpublished English manuscript.]

Quine and Davidson on Language, Thought and Reality. Cambridge: Cambridge University Press (2003), xvi + 311 pp.

This book is also the topic of two special issues:

Teorema 30(1) (2011). In that issue there are two contributions of mine:
Précis of What Is Analytic Philosophy? pp. 13–18.
Replies to my Commentators. pp. 115–242.

Journal for the History of Analytic Philosophy 2 (2013). In that issue:
What Is Analytic Philosophy? p. 1.
Replies to my Commentators. pp. 35–42.

What Is Analytic Philosophy? Cambridge: Cambridge University Press (2008), xii + 292 pp. [translations into French, Spanish, Portuguese, Korean, Farsi, and German.]

Books Edited

Language and Intentionality. Special Issue of *Language and Communication* 21(2) (2001), 103 pp.

Out of Apathy. 30 Years of the British New Left. Ed. with R. Archer et al. London: Verso (1989), viii + 172 pp.

Strawson and Kant. Oxford: Oxford University Press (2003), xii + 259 pp. [Includes an 'Introduction' by the editor, pp. 1–6.]

The Blackwell Companion to Wittgenstein. Ed. with J. Hyman. Marsden: Wiley (2017), pp. xviii + 770.

The Rise of Analytic Philosophy. Oxford: Blackwell (1997), xiv + 95 pp. [Includes an 'Introduction by the Editor', pp. vii–xiv.] [Also as *Ratio* Special Issue 9(3) (1996); Japanese edition: Koyo Shobo 2002.]

Two Dogmas of Empiricism—50 Years After. Ed. with G. Keil and K. Glüer-Pagin. Special Issue of *Grazer Philosophische Studien* 66(2) (2003).

Wittgenstein – A Critical Reader. Oxford: Blackwell (2001), xxvii + 380 pp. [Includes a 'Preface' by the editor, pp. xi–xx.]

Wittgenstein and Analytic Philosophy: Essays in Honour of Peter Hacker. Ed. with J. Hyman. Oxford: Oxford University Press (2009), xxii + 315 pp. [Includes a 'Preface' by the editors, pp. vii–xi.]

Wittgenstein and Quine. Ed. with R. Arrington. London and New York: Routledge (1996), xviii + 286 pp. [Includes an 'Editors' Introduction', pp. xiii–xvii; Second, revised edition: London and New York: London and New York (2003); translation into Farsi.]

Wittgenstein's Philosophical Investigations: Text and Context. Ed. with R. Arrington. London and New York: Routledge (1991), xi + 243 pp. [Includes an 'Editors' Introduction', pp. 1–11.]

Journal Articles

A Cognitivist Approach to Concepts. *Grazer Philosophische Studien* 82 (2011): 111–143.

A Radical Interpretation of Davidson: Reply to Alvarez. *Philosophical Quarterly* 45 (1995): 206–212.

Abusing Use. *Dialectica* 50 (1996): 205–223.

Agency, Intelligence and Reasons in Animals. *Philosophy* 94 (2019): 645–671. https://doi.org/10.1017/S0031819119000275.

Analytic Philosophy and History: A Mismatch? *Mind* 117 (2008): 867–897.

Animal Minds: A Non-Representationalist Approach. *American Philosophical Quarterly* 50 (2013): 213–232.

Animal Minds: Conceptual Problems. *Evolution and Cognition* 5 (1999): 174–188.

Animals, Thoughts and Concepts. *Synthese* 123 (2000): 35–64.

Besser langsam als Wild? Replik auf Markus Wild. *Erwägen, Wissen, Ethik* 23(1) (2012): 56–58.

Cambridge, Jena or Vienna—The Roots of the *Tractatus*. *Ratio*, New Series 5 (1992): 1–23. [reprinted in S. Shanker and D. Kilfoyle (eds.), *Wittgenstein: Critical Assessments, Vol I*. London and New York: Routledge (2001), pp. 31–50).]

Can Animals Act for Reasons? *Inquiry* 52 (2009): 232–255.

Can Animals Judge? *Dialectica* 64 (2010): 11–33.

Concepts, Abilities and Propositions. *Grazer Philosophische Studien* 81 (2010): 115–136.

Concepts, Conceptual Schemes and Grammar. *Philosophia* 37 (2009): 653–668.

Concepts: Where Subjectivism Goes Wrong. *Philosophy* 84 (2009): 5–29. [included in the editor's selection of best articles from 2009.]

Could Anything Be Wrong with Analytic Philosophy? *Grazer Philosophische Studien* 74 (2007): 215–237.

Determinacy of Content: The Hard Problem about Animal Intentionality. *Harvard Review of Philosophy* Online First (2020): 101–120. https://doi.org/10.5840/harvardreview202072631.

Does Ontology Exist? *Philosophy* 77 (2002): 231–256.

Doing Good by Splitting Hairs? Analytic Philosophy and Applied Ethics. *Journal of Applied Philosophy* 28 (2011): 225–240.

Exorcising Grice's Ghost: An Empirical Approach to Studying Intentional Communication in Animals. With S. Townsend, S. Koski, R.W. Byrne, K.E.

Slocombe, B. Bickel, M. Boeckle, I.B. Goncalves, J.M. Burkart, T. Flower, F. Gaunet, T. Gruber, D.A.W.A.M. Jansen, K. Liebal, A. Linke, A. Miklósi, R. Moore, C.P. van Schaik, S. Stoll, A. Vail, B.M. Waller, M. Wild, K. Zuberbühler, M. Manser. *Biological Reviews* 92 (2017): 1427–1433. https://doi.org/10.1111/brv.12289.

Externalism and First Person Authority. With J. Preston. *The Monist* 78 (1995): 515–534.

Forms of Life: Back to Basics. In K. Neumer (ed.), *Das Verstehen des Anderen; Wittgenstein Studies*, 1(1) (2000): 61–84.

Intentionality and Language. *Language and Communication* 21(2) (2001): 105–118.

Kant and Wittgenstein: Philosophy, Necessity and Representation. *International Journal of Philosophical Studies* 5 (1997): 285–305.

Mead and Vygotsky on the Self, Meaning and Internalisation. *Studies in Soviet Thought* 31 (1986): 131–148.

Meaning and Method. *Belgrade Philosophical Annuals* 30 (2017): 7–33. https://doi.org/10.5937/BPA1730007G.

Minds, Brains and Capacities: Situated Cognition and Neo-Aristotelianism. *Frontiers in Psychology* 11 (2020): 1–14. https://doi.org/10.3389/fpsyg.2020.566385.

Necessity and Language: In Defence of Conventionalism. *Philosophical Investigations* 31 (2008): 24–47.

Neural Representationalism. *Facta Philosophica* 5 (2003): 147–171.

Nonsense Made Intelligible. *Erkenntnis* 80 (2015): 111–136.

Persons and Their Bodies. With J. Hyman. *Philosophical Investigations* 17 (1994): 365–379.

Philosophy Rehinged? *International Journal for the Study of Scepticism* 6 (2016): 274–308. https://doi.org/10.1163/22105700-00603012.

Philosophy without Psychology: A Case of Wishful Thinking? *Analiza i Egzystencj* 47 (2019): 5–29. https://doi.org/10.18276/aie.2019.47-01.

Pluralism about Practical Reasons and Reason Explanations. With E. Schmidt. *Philosophical Explorations* (2021). https://doi.org/10.1080/13869795.2021.1908578.

Reasons for Action: Wittgensteinian and Davidsonian Perspectives in Historical and Meta-Philosophical Context. *Nordic Wittgenstein Review* 31 (2014): 7–46.

Reference and the First Person Pronoun. With P.M.S. Hacker. *Language and Communication* 18 (1996): 95–105.

Relativism, Commensurability and Translatability. *Ratio* 20 (2007): 377–402.

Stroud's Rehabilitation of Cartesian Scepticism—A 'Linguistic' Response. *Philosophical Investigations* 13 (1990): 44–64.

The (Limited) Space for Justice in Social Animals. With M. Christen. *Social Justice Research* 25 (2012): 298–326.

The Awful English Language. *Philosophical Papers* 47(1) (2018): 123–154. https://doi.org/10.1080/05568641.2018.1429737.

The Euthanasia Debate in Germany—What's the Fuss? *Journal of Applied Philosophy* 11 (1994): 209–220.

The Indispensability of Translation in Quine and Davidson. *Philosophical Quarterly* 43 (1993): 194–209.

The Object of Philosophy—Tugendhat's Semantical Transformation of Ontology. *Cogito* 8 (1994): 234–241.

Thought, Judgement and Perception. *Grazer Philosophische Studien* 86 (2012): 207–221.

Truth in the *Tractatus*. *Synthese* 148 (2006): 345–368.

Truth without People? *Philosophy* 72 (1997): 85–104.

Was Wittgenstein an Analytic Philosopher? *Metaphilosophy* 35 (2004): 419–444. [French translation: Wittgenstein, philosophe analytique? In E. Rigal (ed.), *Wittgenstein: état des lieux*. Paris: J. Vrin (2008), pp. 330–352.]

What are Concepts? *Conceptus: Zeitschrift für Philosophie* 39 (2010): 7–39.

What Is a Theory of Meaning? Just When You Thought Conceptual Analysis Was Dead. *Cahiers Ferdinand de Saussure* 65 (2012): 51–79.

Wie kam die Bedeutung zur Regel? *Deutsche Zeitschrift für Philosophie* 48 (2000): 429–447.

Wie wichtig ist Erkenntnistheorie? *Zeitschrift für Philosophische Forschung* 56 (2002): 96–117.

Wittgensteinian Anti-Anti Realism: One 'Anti' Too Many? *Ethical Perspectives* 22 (2015): 99–129. https://doi.org/10.2143/EP.22.1.3073459. [reprinted in R. Amesbury and H. von Sass (eds.), *Ethics after Wittgenstein*. London: Bloomsbury (2021), pp. 99–127.]

Contributions to Anthologies

A Correspondence Theory of Truth? In P. Frascolla (ed.), *Tractatus logico-philosophicus: Sources, Themes, Perspectives, Proceedings of the International Workshop – Lagopesole, 25th–26th October 2000. Annali della Facoltà di Lettere e Filosofia dell'Università degli Studi della Basilicata, n. 11*. Potenza: Graphis (2002), pp. 52–74.

All Kinds of Nonsense. In E. Ammereller and E. Fischer (eds.), *Wittgenstein at Work*. London: Routledge (2004), pp. 221–245.

Analytic Philosophy and Idealism. In J. Padilla-Gálvez (ed.), *Idealismus und sprachanalytische Philosophie*. Frankfurt: Peter Lang (2007), pp. 91–112.

Analytic Philosophy: Wittgenstein and After. In D. Moran (ed.), *A Companion to Twentieth-Century Philosophy*. London: Routledge (2008), pp. 76–117.

Animal Belief and Rationality. In K. Andrews and J. Beck (eds.), *The Routledge Handbook of Philosophy and Animal Minds*. London: Routledge (2017), pp. 89–99.

Animal Consciousness – A Limit of Language? In H. Appelquist (ed.), *Wittgenstein and the Limits of Language*. Abingdon: Routledge (2020), pp. 197–222.

Animal Minds. In B. McLaughlin (ed.), *Philosophy of Mind. Macmillan Interdisciplinary*. Farmington Hills, MI: Macmillan Reference (2016), pp. 327–351.

Animal Minds: Philosophical and Scientific Aspects. In T.P. Racine and K.L. Slaney (eds.), *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*. Basingstoke: Palgrave (2013), pp. 130–152.

Animals: Agency, Reasons and Reasoning. In J. Nida-Rümelin and E. Özmen (eds.), *Welt der Gründe. Proceedings XXII. Deutscher Kongress für Philosophie*. Hamburg: Meiner (2012), pp. 900–913.

Apriority, Analyticity and Language. In N. Kompa, C. Nimtz and C. Suhm (eds.), *The A Priori and Its Role in Philosophy*. Paderborn: Mentis (2009), pp. 119–136.

Aristotle on the Anthropological Difference and Animal Minds. In G. Keil and N. Kreft (eds.), *Aristotle's Anthropology*. Cambridge: Cambridge University Press (2019), pp. 140–160.

Aspect-Perception, Perception and Animals. In G. Kemp and G. Mras (eds.), *Wollheim and Wittgenstein: Seeing-as/in and Art*. Abingdon: Routledge (2016), pp. 77–100.

Ayer's Verificationism: Dead as a Dodo? In A.T. Tuboly (ed.), *The Historical and Philosophical Significance of Ayer's 'Language, Truth and Logic'*. Cham: Palgrave (2021), pp. 251–278.

Begriffliche Probleme und das Problem der Begriffe. In D. Perler and M. Wild (eds.), *Der Geist der Tiere*. Frankfurt: Suhrkamp (2005), pp. 153–187.

'Clarity' Is Not Enough. In R. Haller and K. Puhl (eds.), *Wittgenstein and the Future of Philosophy. Proceedings of the 24th International Wittgenstein Symposium*. Vienna: Hölder-Pichler-Tempsky (2002), pp. 81–98.

Concepts and Experience: A Non-Representationalist Approach. In C. Demmerling and D. Schröder (eds.), *Concepts in Thought, Action, and Emotion*. Abingdon: Routledge (2021), pp. 21–41.

Concepts: Representations or Abilities? In E. Di Nucci and C. McHugh (eds.), *Content, Consciousness, and Perception: Essays in Contemporary Philosophy of Mind*. Cambridge: Cambridge Scholars Press (2006), pp. 37–61.

Concepts; Between the Subjective and the Objective. In J. Cottingham and P.M.S. Hacker (eds.), *Mind, Method and Morality: Essays in Honour of Anthony Kenny*. Oxford: Oxford University Press (2010), pp. 306–329.

Critical Note: Language and Learning from the 4E Perspective. In A. Newen, L. de Bruin and S. Gallagher (eds.), *The Oxford Handbook of 4E Cognition*. Oxford: Oxford University Press (2018), pp. 707–716.

Der Neukantianismus und die Analytische Philosophie. In R. Alexy, L.H. Meyer, S.L. Paulsen and G. Sprenger (eds.), *Neukantianismus und Rechtsphilosophie*. Baden-Baden: Nomos (2002), pp. 499–513.

Die schreckliche englische Sprache. In W. Lütterfelds and T. Mohrs (eds.), *Die Welt ist meine Welt—Kulturelle Aspekte der Globalisierung*. Frankfurt: Peter Lang (2004), pp. 49–67.

Does Language Require Conventions? In P. Frascolla, D. Marconi and A. Voltolini (eds.), *Wittgenstein: Mind, Meaning and Metaphilosophy*. London: Palgrave (2010), pp. 85–112.

Eine ganze Wolke von Philosophie kondensiert zu einem Tröpfchen Sprachlehre. In E. von Savigny and O. Scholz (eds.), *Wittgenstein über die Seele*. Frankfurt: Suhrkamp (1995), pp. 233–252.

Evolutionäre Erkenntnistheorie. In P. Schmid-Hempel and H.U. Reyer (eds.), *Evolution*. Zürich: vdf (2011), pp. 235–249.

Filosofia, estetica e critica culturale. In E. Caldarola, D. Quattrocchi and G. Tomasi (eds.), *Wittgenstein, l'estetica e le arti*. Roma: Carroci (2013), pp. 137–155.

Flüchtige Gedanken über Nachdenklichkeit. In C. Abbt, D. Schoeller and H. von Sass (eds.), *Nachdenklichkeit*. Zürich: Diaphanes (2020), pp. 59–69.

Imposters, Bunglers and Relativists. In S. Peters, M. Biddiss and I. Roe (eds.), *The Humanities at the Millennium*. Tübingen: Francke Verlag (2000), pp. 249–269.

Impure Conceptual Analysis. In S. Overgaard and G. d'Oro (eds.), *The Cambridge Companion to Philosophical Methodology*. Cambridge: Cambridge University Press (2017), pp. 83–107.

Insignificant Others: The Mutual Prejudices of Anglophone and Germanophone Philosophers. In C. Brown and T. Seidel (eds.), *Cultural Negotiations*. Tübingen: Francke Verlag (1998), pp. 83–98.

Investigations §128: Theses in Philosophy and Undogmatic Procedure. In R. Arrington and H. Glock (eds.), *Wittgenstein's Philosophical Investigations*. London and New York: Routledge (1991), pp. 69–88. [reprinted in S. Shanker and D. Kilfoyle (eds.), *Ludwig Wittgenstein: Critical Assessments, Vol. II*. London and New York: Routledge (2001), pp. 52–67.]

Judgement and Concepts in Animals. In J. Burgos and E. Ribes-Iñesta (eds.), *The Brain Behavior Nexus: Conceptual Issues. Proceedings of the 10th Biannual Symposium on the Science of Behaviour*. Guadalajara: University of Guadalajara Press (2008), pp. 149–177.

Judgement and Truth in the Early Wittgenstein. In M. Textor (ed.), *Judgement and Truth in Early Analytic Philosophy*. Basingstoke: Palgrave (2013), pp. 242–270.

Knowledge, Certainty and Scepticism: in Moore's Defence. In D. Moyal-Sharrock (ed.), *The Third Wittgenstein*. Aldershot: Ashgate (2004), pp. 63–78.

Ludwig Wittgenstein – Sprache, Bedeutung und Gebrauch. In A. Beckermann and D. Perler (eds.), *Klassiker der Philosophie Heute*. Stuttgart: Reclam (2004), pp. 601–622.

Ludwig Wittgenstein: *Tractatus Logico-Philosophicus*. In J.J.E. Gracia, G.M. Reichenberg and B.N. Schumacher (eds.), *The Classics of Western Philosophy*. Oxford: Blackwell (2002), pp. 474–485.

Mathematik und Begriffsbildung. With K. Büttner. In J. Bromand and B. Reichardt (eds.), *Wittgenstein und die Philosophie der Mathematik*. Paderborn: Mentis (2018), pp. 227–246.

Meaning, Rules and Conventions. In D. Levy and E. Zamuner (eds.), *Wittgenstein's Enduring Arguments*. London: Routledge (2008), pp. 156–178.

Mental Capacities and Animal Ethics. In K. Petrus and M. Wild (eds.), *Animal Minds and Animal Ethics*. Bielefeld: Transcript (2013), pp. 113–146.

Necessary Truth and Grammatical Propositions. In J. Padilla-Gálvez (ed.), *Phenomenology as Grammar*. Frankfurt: Ontos (2008), pp. 63–76.

Necessity and Normativity. In H. Sluga and D. Stern (eds.), *The Cambridge Companion to Wittgenstein*. Cambridge and New York: Cambridge University Press (1996), pp. 198–225 [2nd edn. 2018, pp. 180–208].

Necessity, A Priority and Analyticity. In D. Whiting (ed.), *The Later Wittgenstein on Language*. Basingstoke: Palgrave (2010), pp. 133–147.

Nelson und die Analytische Philosophie. In A. Berger, G. Raupach-Strey and J. Schroth (eds.), *Leonard Nelson: ein früher Denker der Analytischen Philosophie?* Münster: Lit-Verlag (2011), pp. 39–70.

Neo-Kantianism and Analytic Philosophy. In N. de Warren and A. Staiti (eds.), *New Approaches to Neo-Kantianism*. Cambridge: Cambridge University Press (2015), pp. 59–81.

Non-Human Knowledge and Non-Human Agency. In S. Tolksdorf (ed.), *Conceptions of Knowledge*. New York: de Gruyter (2012), pp. 521–551.

Objectivism and Causalism about Reasons for Action. With E. Schmidt. In G. Schumann (ed.), *Explanation in Action Theory and Historiography: Causal and Teleological Approaches*. Routledge Studies in Contemporary Philosophy. Abingdon: Routledge (2019), pp. 124–145.

On Safari with Wittgenstein, Quine and Davidson. In R. Arrington and H. Glock (eds.), *Wittgenstein and Quine*. London: Routledge (1996; 2nd edition 2004), pp. 144–173.

Ontologie—gibt's das wirklich? In C. Nimtz et al. (eds.), *Argument und Analyse. Proceedings of GAP 4*. Paderborn: Mentis (2001), pp. 436–447.

P.F. Strawson: Ordinary Language Philosophy and Descriptive Metaphysics. In A. Preston (ed.), *Analytic Philosophy: An Interpretive History*. London: Routledge (2017), pp. 214–228.

Perception and the Brain. In T. Fischer-Seidel, S. Peters and A. Potts (eds.), *Perception and the Senses—Sinneswahrnehmung*. Tübingen: Francke (2003), pp. 39–50.

Perspectives on Wittgenstein: An Intermittently Opinionated Survey. In G. Kahane, E. Kanterian and O. Kuusela (eds.), *Wittgenstein's Interpreters*. Oxford: Blackwell (2007), pp. 37–65.

Philosophical Investigations: Principles of Interpretation. In W. Brandl and R. Haller (eds.), *Wittgenstein—A Reevaluation. Proceedings of the 14th International Wittgenstein-Symposium*. Vienna: Hölder-Pichler-Tempsky (1990), pp. 152–162.

Philosophy and Philosophical Method. In H.J. Glock and J. Hyman (eds.), *The Blackwell Companion to Wittgenstein*. Marsden: Wiley (2017), pp. 231–251.

Philosophy of Language. In M. Forster and K. Gjesdal (eds.), *The Oxford Handbook of Nineteenth-Century German Philosophy*. New York: Oxford University Press (2015), pp. 371–397.

Philosophy, Thought and Language. In J. Preston (ed.), *Thought and Language: Proceedings of the Royal Institute of Philosophy Conference*. Cambridge: Cambridge University Press (1997), pp. 151–169.

Propositional Attitudes, Intentional Contents and Other Representationalist Myths. In A. Coliva, D. Moyal-Sharrock and V. Munz (eds.), *Mind, Language and Action*. New York: de Gruyter (2014), pp. 512–537.

Quine and Davidson. In E. LePore and K. Ludwig (eds.), *The Blackwell Companion to Davidson*. New York: Wiley (2013), pp. 567–587.

Radical Translation and Conceptual Relativism. *The European Legacy Vol. 2* (1997), pp. 603–608.

Ramsey and Wittgenstein: Mutual Influences. In M.J. Frápolli (ed.), *F. P. Ramsey: Critical Reassessments*. London and New York: Continuum Studies in British Philosophy (2005), pp. 41–68.

Schopenhauer and Wittgenstein: Representation as Language and Will. In C. Janaway (ed.), *The Cambridge Companion to Schopenhauer*. Cambridge and New York: Cambridge University Press (1999), pp. 422–458.

Semantics: Why Rules Ought to Matter. In O.L. Koreň and V. Kolman (eds.), *Rules Matter*. Abingdon: Routledge (2018), pp. 63–80.

Sense and Meaning in Frege and the *Tractatus*. In G. Oliveri (ed.), *From the Tractatus to the Tractatus and Other Essays*. Frankfurt a.M.: Peter Lang (2000), pp. 53–68.

Sprache und Geist. In A. Kolmer, M. Meyer and E. Stark (eds.), *Sprache(n) Verstehen*. Zürich: vdf (2014), pp. 85–100.

Strawson and Analytic Kantianism. In H.J. Glock (ed.), *Strawson and Kant*. Oxford: Clarendon Press (2003), pp. 15–42.

Strawson and Non-Revisionary Naturalism. In M. De Caro and D. Macarthur (eds.), *The Routledge Handbook of Liberal Naturalism*. Abingdon: Routledge (2022), chapter 36.

Strawson's Descriptive Metaphysics. In L. Haaparanta and H. Koskinnen (eds.), *Categories of Being*. New York: Oxford University Press (2012), pp. 391–419.

The Analytic/Continental Disagreement. In M.C. Amoretti and M. Vignolo (eds.), *Disaccordo. Annuario della Società Italiana di Filosofia Analitica (SIFA) 2012*. Milano: Mimesis (2012), pp. 149–172.

The Anthropological Difference. In C. Sandis and M.J. Cain (eds.), *Human Nature. Royal Institute of Philosophy Supplement: 70*. Cambridge: Cambridge University Press (2012), pp. 105–131.

The Bounds of Sense and the Rules of Grammar. In W. Leinfellner and F. Wuketits (eds.), *The Contemporary Task of Philosophy*. Vienna: Hölder-Pilcher-Tempsky (1986), pp. 516–518.

The Development of Wittgenstein's Philosophy. In. H.J. Glock (ed.), *Wittgenstein: A Critical Reader*. Oxford: Blackwell (2001), pp. 1–25.

The Influence of Wittgenstein on American Philosophy. In C. Misak (ed.), *The Oxford Handbook of American Philosophy*. New York: Oxford University Press (2008), pp. 375–402.

The Linguistic Doctrine Revisited. In G. Keil and K. Glüer-Pagin (eds.), *Two Dogmas of Empiricism—50 Years After. Special Issue of Grazer Philosophische Studien* 66(2) (2003), pp. 19–46.

The Normativity of Meaning made Simple. In A. Beckermann and C. Nimtz (eds.), *Philosophy and Science. Proceedings of GAP 5*. Paderborn: Mentis (2005), pp. 219–241.

The Normativity of Meaning Revisited. In K. Bayertz and N. Roughley (eds.), *The Normative Animal*. Oxford: Oxford University Press (2019), pp. 295–318.

The Owl of Minerva: Is Analytic Philosophy Moribund? In E. Reck (ed.), *The Historical Turn in Analytic Philosophy*. London: Palgrave (2013), pp. 326–347.

The Relation between Quine and Davidson. In G. Harman and E. Lepore (eds.), *The Blackwell Companion to Quine*. New York: Wiley (2014), pp. 526–551.

Thought, Language and Animals. In M. Kober (ed.), *Deepening Our Understanding of Wittgenstein. Special Issue of Grazer Philosophische Studien 71* (2006), pp. 139–160.

Tractatus Logico-Philosophicus. In J. Shand (ed.), *Central Works of Philosophy Vol. 4*. Chesham: Acumen (2006), pp. 71–91.

Unverständlichkeit verständlich machen. In T. Petraschka et al. (eds.), *Fiktion, Wahrheit, Interpretation*. Paderborn: Mentis (2013), pp. 220–239.

Vorsprung durch Logik: The German Analytic Tradition. In A. O'Hear (ed.), *German Philosophy since Kant. Lectures of the Royal Institute of Philosophy*. Cambridge: Cambridge University Press (1999), pp. 137–166.

Wann ist ein Tier ein Tier? In M. Fehlmann, M. Michel and R. Niederhauser (eds.), *Tierisch!* Zürich: vdf (2016), pp. 13–22.

Was ist Denken? (Replik auf S. Tietz). In H.J. Rüegger et al. (eds.), *Abschied vom Seelischen*. Zürich: vdf (2013), pp. 231–234.

What Have the Historians Ever Done for Us? In M. van Ackeren (ed.), *Philosophy and the Historical Perspective*. Oxford: Oxford University Press (2018), pp. 18–35.

What Is Meaning? A Wittgensteinian Answer to an UnWittgensteinian Question. In S. Greve and J. Conant (eds.), *Wittgenstein on Objectivity, Intuition, and Meaning*. New York: Cambridge University Press (2019), pp. 185–210.

Wittgenstein and 'Ordinary Language Philosophy'. With J. Kalhat. In A. Matar (ed.), *Understanding Wittgenstein, Understanding Modernism*. London: Bloomsbury (2016), pp. 55–74.

Wittgenstein and Davidson on Animal Minds. In C. Verheggen (ed.), *Wittgenstein and Davidson on Thought, Language and Action*. Cambridge: Cambridge University Press (2018), pp. 159–185.

Wittgenstein and History. In Alois Pichler and Simo Säätelä (eds.), *Wittgenstein: The Philosopher and His* Works. Bergen: Wittgenstein Archives at the University of Bergen (2005), pp. 177–204. [reprinted in a second edition published by Ontos, 2006.]

Wittgenstein and Quine: Mind, Language and Behaviour. In S. Schroeder (ed.), *Wittgenstein and Contemporary Philosophy of* Mind. Basingstoke: Palgrave (2001), pp. 3–23.

Wittgenstein and Reason. In J. Klagge (ed.), *Wittgenstein: Biography and Philosophy*. New York: Cambridge University Press (2001), pp. 195–220.

Wittgenstein on Concepts. In A.M. Ahmed (ed.), *Wittgenstein's "Philosophical Investigations": A Critical Guide*. Cambridge: Cambridge University Press (2010), pp. 88–108.

Wittgenstein on Truth. In W. Löffler and P. Weingartner (eds.), *Knowledge and Belief*. Vienna: Hölder-Pichler-Tempsky (2004), pp. 328–346.

Wittgenstein vs. Quine on Logical Necessity. In S. Teghrarian (ed.), *Wittgenstein and Contemporary Philosophy*. Bristol: Thoemmes Press (1994), pp. 185–222.

Wittgenstein's Conventionalism. In A. Coliva and E. Picardi (eds.), *Wittgenstein Today*. Padova: Il Poligrafo (2004), pp. 143–167.

Wittgenstein's Later Philosophy. In M. Beaney (ed.), *The Oxford Handbook of the History of Analytic Philosophy*. New York: Oxford University Press (2013), pp. 572–593.

Wittgenstein's Rain in the Philosophical Desert. In H. Price and C. Misak (eds.), *Proceedings of the British Academy 210*, *The Practical Turn: Pragmatism in Britain in the Long 20th Century*. Oxford: Oxford University Press (2017), pp. 131–148.

Wittgensteins letzter Wille. In E. von Savigny (ed.), *Klassiker Auslegen: Wittgensteins Philosophische Untersuchungen*. Leipzig: Akademie Verlag (1998), pp. 215–237.

Articles in Works of Reference and Handbooks

Analytische Philosophie (I): W.V. Quine. In J. Urbich (ed.), *Handbuch Ontologie*. Stuttgart: Metzler (2020), pp. 236–242.

Analytische Philosophie. In S. Jordan and C. Nimtz (eds.), *Lexikon Philosophie*. Stuttgart: Reclam (2009), pp. 26–29.

Animal Agency. In C. Sandis and T. O'Connor (eds.), *A Companion to the Philosophy of Action*. Oxford: Wiley-Blackwell (2010), pp. 384–392.

Frege, Gottlob. In R.L. Arrington (ed.), *A Companion to the Philosophers*. Oxford: Blackwell (1999), pp. 253–260. [reprinted in R.L. Arrington (ed.), *The World's Great Philosophers*. London: Blackwell (2003), pp. 82–91.]

Geist der Tiere. In R. Borgards (ed.), *Tiere. Ein interdisziplinäres Handbuch*. Stuttgart: Metzler (2016), pp. 60–78.

Gottlob Frege. In S. Jordan and B. Mojsisch (eds.), *Philosophenlexikon*. Stuttgart: Reclam (2009), pp. 120–121.

Habermas, Jürgen. In J. Dancy and E. Sosa (eds.), *The Blackwell Companion to Epistemology*. Oxford: Blackwell (1992), pp. 166–167.

Habermas, Jürgen. In J. Sandford (ed.), *The Routledge Companion to German Culture*. London: Routledge (1999), pp. 272–274.

Intentionalität. In A. Ferrari and K. Petrus (eds.), *Lexikon der Mensch/Tier-Beziehungen*. Bielefeld: Transcript (2015), pp. 170–173.

Kriterium. With Pit Genot. In S. Majetschak and A. Weiberg (eds.), *Wittgenstein-Handbuch*. Stuttgart: Metzler (2022), pp. 282–284.

Language, Philosophy of. In A. Grayling, N. Goulder, and A. Pyle (eds.), *The Continuum Encyclopedia of British Philosophy*. Vol. 3. London: Thoemmes Continuum (2006), pp. 1807–1812.

Ludwig Wittgenstein. In S. Jordan and B. Mojsisch (eds.), *Philosophenlexikon*. Stuttgart: Reclam (2009), pp. 32–327.

Meaning and Rule-Following. In J.D. Wright (ed.), *International Encyclopedia of Social and Behavioral Sciences*. 2nd edition. Vol. 14. Amsterdam: Elsevier (2015), pp. 841–849. Ontology. In A. Grayling, N. Goulder, and A. Pyle (eds.), *The Continuum Encyclopedia of British Philosophy*. Vol. 3. London: Thoemmes Continuum (2006), pp. 2378–2380.

Ordinary Language Philosophy. In S. Majetschak and A. Weiberg (eds.), *Wittgenstein-Handbuch*. Stuttgart: Metzler (2022), pp. 455–460.

Peter Strawson. In S. Jordan and B. Mojsisch (eds.), *Philosophenlexikon*. Stuttgart: Reclam (2009), p. 304.

Philosophy. In J. Sandford (ed.), *The Routledge Companion to German Culture*. London: Routledge (1999), pp. 477–480.

Sprache. In A. Ferrari and K. Petrus (eds.), *Lexikon der Mensch/Tier-Beziehungen*. Bielefeld: Transcript (2015), pp. 327–329.

Strawson, Peter Frederick. In R.L. Arrington (ed.), *A Companion to the Philosophers*. Oxford: Blackwell (1999), pp. 524–528.

Tiere und Denken. In J.S. Ach and D. Borchers (eds.), *Handbuch Tierethik*. Stuttgart: Metzler (2018), pp. 52–56.

Verhalten. With E. Hämmerli. In S. Majetschak and A. Weiberg (eds.), *Wittgenstein-Handbuch*. Stuttgart: Metzler (2022), pp. 407–411.

Reviews

T. Baldwin, Contemporary Philosophy. *British Journal for the History of Philosophy* 12 (2004): 564–569.

J. Chase and J. Reynolds, Analytic versus Continental: Arguments on the Methods and Value of Philosophy. *Australasian Journal of Philosophy* 90 (2012): 398–402.

B. de Mesel, The Later Wittgenstein and Moral Philosophy. *Ratio* (2021). https://doi.org/10.1111/rati.12319.

C. Diamond, The Realistic Spirit & C. Barrett, Wittgenstein on Ethics and Religious Belief. *Times Higher Education Supplement* (May 1992): 25.

'Dummett on the Roots of Analytical Philosophy', review of M. Dummett, Ursprünge der Analytischen Philosophie. *Mind* 98 (1989): 646–649.

M. Dummett, The Nature and Future of Philosophy. *Notre Dame Philosophical Reviews* 4 (2012). http://ndpr.nd.edu/news/29165-the-nature-and-future-of-philosophy/.

P. Frascolla, Wittgenstein's Philosophy of Mathematics. *Philosophical Quarterly* 47 (1997): 552–555.

J. Genova, Wittgenstein: A Way of Seeing. *Mind* 111 (2002): 105–109.

P.M.S. Hacker, Wittgenstein: Mind and Will. *Times Higher Education Supplement* (10 October 1997): 31.

P.M.S. Hacker, Wittgenstein's Place in Twentieth Century Analytic Philosophy and R. Monk and A. Palmer (eds.), Bertrand Russell and the Origins of Analytic Philosophy. *European Journal of Philosophy* 7 (1999): 361–367.

P. Johnston, Wittgenstein and Ethics. *Cogito* 6 (1992): 181–182.

E.M. Lange, Wittgenstein und Schopenhauer. *Philosophical Investigations* 16 (1993): 89–93.

B. McGuinness and G.H. von Wright (eds.), Ludwig Wittgenstein: Cambridge Letters. *International Journal of Philosophical Studies* 7 (1999): 132–135.

M. Nedo (ed.), Wittgenstein Wiener Ausgabe. *Times Literary Supplement* (23 June 1995): 9–10.

R. Nozick, Invariances: The Structure of the Objective World. *Nature* 417 (27 June, 2002): 900–901.

H. Wettstein, The Magic Prism. *Notre Dame Philosophical Reviews* 12 (2005). http://ndpr.nd.edu/news/24930-the-magic-prism-an-essay-in-the-philosophy-of-language/.

Critical Notices (Review Articles)

Critical Discussion: *Wittgensteins Philosophische Untersuchungen. Ein Kommentar* by E. von Savigny. *Erkenntnis* 36 (1992): 117–128.

Frege By A. Kenny. *Grazer Philosophische Studien* 52 (1996/1997): 205–223.

From Armchair to Reality? (T. Williamson's *Philosophy of Philosophy*). *Ratio* 23 (2010): 339–348.

Logic and Natural Language by H. Ben-Yami. *Language and Communication* 27 (2007): 28–40.

Subjective, Intersubjective, Objective by D. Davidson. *Philosophical Investigations* 26 (2003): 348–360.

Articles in Non-Academic Journals

Am Ende des Lateins? *Horizonte* 97 (2013): 9.

Durch Open Access geht Bildung verloren. *Tagesanzeiger* (03.08.2009): 7.

Ein Buch fürs Leben: Denkspiele vom Reissbrett. *UZH Magazin* 2/17 (2017): 8.

From Where I Sit – Storms in a Swiss Teacup. *Times Higher Education* (16.07.2009). http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=407361.

Kants Erkenntnistheorie Heute. *Schweizer Monatshefte* May/June 2004: 31–36.
KI und Intelligenz: Computer sind nicht neugierig. *UZH Magazin* 1/20 (2020): 32–37.
Letter from Switzerland. *The Philosopher's Magazine* (2012) 1st Quarter: 47–50.
Probleme mit Deutschen an Schweizer Unis sind hausgemacht. *Sonntag* (03.02.2008): 17.
Stimmt es, dass Tiere auch denken können. *unijournal* Universty of Zurich (07.05.2007): 20.
Obituary: Oswald Hanfling. *The Guardian* (29.11.2005).
Words and Things. *Prospect* (April 1999): 26–30.

Interviews etc.

Ameisen kennen keine Mathematik. With Rolf Pfeiffer. *Unimagazin* 23(1) (2014): 30–34.
Interview with Hans-Johann Glock. *Philosophical Inquiries* 1 (2018) Special Issue. History of Late Analytic Philosophy. G. Bonardi and P. Tripodi (eds.), pp. 26–29.
Preface to this edition. In Ernst Tugendhat, *Traditional and Analytical Philosophy*, trl. P. Gorner. Cambridge: Cambridge University Press (2016), pp. ix–x.
Was ist mit dem Wissen der Tiere? *Information Philosophie* 11 (2011): 38–45.
Wittgensteins Wahrheitsauffassung: ein Gespräch mit dem Wittgenstein-Experten Hans-Johann Glock. *Information Philosophie* 4 (2003): 125–128.

Index

Note: Page numbers followed by "n" denote endnotes.